Deep Sequencing and Functional Analyses Identify a Role of Fusobacterium Species in Colorectal Tumorigenesis

Deep sequencing and functional analyses identify a role of *Fusobacterium* species in colorectal tumorigenesis

A dissertation presented

by

Aleksandar David Kostic

to

The Division of Medical Sciences

in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in the subject of

Genetics

Harvard University
Cambridge, Massachusetts

April 2013

Dissertation Advisor: Dr. Matthew Meyerson Aleksandar David Kostic

Deep sequencing and functional analyses identify a role of *Fusobacterium* species in colorectal tumorigenesis

Abstract

The tumor microenvironment is a complex community consisting of neoplastic cells, surrounding stromal cells, a broad array of immune cells, and a microbiota. By sheer numbers, the microbiota has its greatest manifestation in colorectal cancer (CRC) because the colon contains up to 100 trillion bacteria, outnumbering human cells by a factor of 10 and encoding a gene-content that is 100-fold larger than that of the human genome. Indeed, previous studies using germ-free mice in a variety of genetic backgrounds have demonstrated that the microbiota can impact colorectal tumorigenesis. In addition, specific strains of enterotoxigenic bacteria have been shown to promote colitis-associated cancer in mice. Here, we explore the composition of the tissue-associated microbiota in human CRC and evaluate the role of tumor-enriched microbes in potentiating colorectal tumorigenesis in mice.

Advances in DNA sequencing technology have fueled a renaissance in the microbiome field. Deep sequencing metagenomics enables rapid, culture-independent characterization of a microbial community. We present PathSeq, a highly scalable software tool that performs computational subtraction on high-throughput sequencing data to identify nonhuman nucleic acids. PathSeq makes it possible to analyze sequence datasets as large as human whole-genomes for the purpose of metagenomics and also to discover previously unsequenced microorganisms. We used PathSeq to characterize the composition of the microbiota in human CRC using whole-

genome sequencing on nine tumor/normal pairs and 16S rDNA sequencing on an additional 95 pairs. The genus *Fusobacterium* was highly enriched in tumors, while the Bacteroidetes and Firmicutes phyla were depleted.

We show that in the $Apc^{Min/+}$ mouse model of intestinal tumorigenesis, *Fusobacterium nucleatum* increases tumor multiplicity, selectively recruits tumor-infiltrating myeloid cells, and is associated with a pro-inflammatory expression signature that is shared with human fusobacteria-positive colorectal carcinomas. We find that *Fusobacterium* spp. are enriched in human colonic adenomas relative to surrounding tissues and fusobacterial abundance is increased in stool samples from patients with colorectal adenomas and carcinomas, compared to healthy subjects. Collectively, these data support that fusobacteria may be involved in early stages of intestinal tumorigenesis and, through recruitment of tumor-infiltrating immune cells, may generate a pro-inflammatory tissue microenvironment conducive to colorectal neoplasia progression.

Table of Contents

Preface……..i
Abstract……iii
Table of Contents……...……v
Acknowledgements……vii
List of Figures……...ix
List of Tables……...……xi
List of Abbreviations……...xii

Introduction: **Microbes and Inflammation in Cancer**……1
Overview……..……..2
The Human Gastrointestinal Microbiota……...3
Overview……..3
The Intestinal Ecosystem: The Flora……...4
The Intestinal Ecosystem: The Host……6
Colorectal Cancer……..6
The Role of Inflammation in Cancer……8
The Role of Inflammatory Responses to the Microbiota in Cancer……9
MyD88……9
TLRs……..11
NLRs and the Inflammasome……12
Defensins……14
NF-κB Activation and the IL-6—STAT3 Axis……14
IL-23……..15
COX-2……16
Mouse Models of Immunodeficiency in Colitis and Cancer……16
The Role of the Gut Microbiota in Colorectal Cancer……17
Associations between Single Species and Cancer……17
Evidence for a Role of the Microbiota in CRC from Gnotobiotic Mice……...20
Microbial Products that Contribute to Tumorigenesis……...22
β-glucuronidase……..23
7-α-dehydroxylase……24
Reactive Oxygen Intermediates……24
N-nitroso Compounds, Heterocyclic Amines, and Other Products……...25
Bacteria that Show Protective Effects in CRC……..26
Probiotics……...27
Intestinal Barrier Defects and Microbial Infiltration in CRC……28
Thesis Summary……..29
References……30

Chapter 1: **PathSeq: software to identify or discover microbes by deep sequencing of human tissue**……..49
Abstract……..50
Introduction……50
Results and Discussion……..52

Materials and Methods……...58
References……..65

Chapter 2: **Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma**……67
Abstract……..68
Introduction……68
Results……69
Discussion……..76
Materials and Methods……...78
References……..83

Chapter 3: ***Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor immune microenvironment**……86
Abstract……..87
Introduction……87
Results……88
Fusobacterium nucleatum promotes intestinal tumorigenesis……...88
F. nucleatum selectively expands myeloid-derived immune cells……91
A *Fusobacterium*-associated human colorectal cancer gene signature shared and validated in mice……94
Fusobacteria are enriched in colonic adenomas and in stools samples from patients with adenomas and colorectal carcinomas……...97
Discussion……..99
Materials and Methods……102
References……108

Conclusion: **Perspectives on the Study of the Microbiota in Colorectal Cancer**……114

Appendix 1……...117
Supplementary Materials……117

Acknowledgements

Firstly, I would like to thank my dissertation advisor, Matthew Meyerson, for his tremendous support throughout my years in his lab. Though at first I had the intention of working on cancer genetics, he somehow changed my mind entirely within 15 minutes of meeting with him. His energy and enthusiasm for science, and for pathogen discovery in particular, is…infectious. I spent my first year learning bioinformatics, programming, and statistics rather than pumping out results, but Matthew was incredibly supportive during this time. As I developed PathSeq he gave me space to explore and be creative, but when it came time to publish he kept me grounded and held me to account. During my later years in the lab he entrusted me with a lot of independence but at the same time carefully ensured that I was making good progress. Matthew is an exceptional mentor and a role model for me as I continue my career in science.

As my project took me away from computational biology and into the entirely new realm of immunology and mouse work, I had the incredible privilege of receiving mentorship from Wendy Garrett. Wendy is a dedicated scientist who is available to her trainees at all hours of the day, literally. She has opened my imagination to the great potential that the fields of the microbiome and immunology hold, and has also given me insight and first-hand experience in clinical oncology. She did her best to train me in mouse work and humored me that I would eventually get comfortable with sacking mice…but unfortunately that hasn't yet happened.

In my early years as a graduate student, Akin Ojesina's great enthusiasm and out-of-the-box ideas were formative for me and got me interested in the microbiome field in the first place. My productivity on the computational end was due in large part to the great support I received from Chandra Pedamallu, a tireless and dedicated computational biologist. I am indebted to Gad

Getz, Craig Mermel, Roel Verhaak, and Michael Berger for their early support when I was getting started with computational work, and Dirk Gevers and Curtis Huttenhower for teaching me the ropes of metagenomics. I also thank Joonil Jung and Ami Bhatt for their valuable scientific and career advisement over the years.

I have many people to thank for fruitful, and sometimes less fruitful but amusing, discussions and collaborations during my time in the Meyerson lab and at the Broad Institute: Adam Bass, Nicolas Stransky, Alexis Ramos, Scott Carter, Trevor Pugh, Marcin Imielinski, Shantanu Banerji, Jordi Barretina, Derek Chiang, Amit Dutt, Heidi Greulich, Josh Francis, Peter Hammerman, Luc De Waal, Rachel Liao, Tanaz Sharifnia, Kumiko Tanaka, Tzu-Hsiu Chen, and Fujiko Duke. I owe a lot to the members of the Garrett Lab not only for their companionship and for making me feel at home in their lab over the years, but for their training in immunology, microbiology, and mouse techniques: Eunyoung Chung, Patrick Smith, Michael Howitt, Patrick Veiga, Sonia Ballal, Leslie Wardwell, Michelle Rooks, Monia Michaud, Carey Anne Gallini, and Lauren Robertson.

I would like to thank the members of my Dissertation Advisory Committee: Gary Ruvkun, Shamil Sunyaev, and Soumya Raychaudhuri. They have been a helpful presence at numerous junctures during my thesis work, and I have benefited greatly from formal and informal meetings with them.

Last but certainly not least, I would like to thank my friends both here in Boston and back home in Toronto for making life outside of the lab interesting, and my caring family for their encouragement through grad school. Finally, I couldn't have done this without the loving support I have received from Diane Shao, who has been there to help me get over the defeats and celebrate even the smallest victories through most of my journey as a graduate student.

List of Figures

Figure 1-1. The PathSeq workflow……53

Figure 1-2. PathSeq performance on simulated and experimental sequence data……...55

Figure 2-1. Whole-genome sequencing analysis of the colorectal cancer microbiome……71

Figure 2-2. 16S rDNA sequencing analysis of the colorectal cancer microbiome……72

Figure 2-3. Fluorescence *in situ* hybridization (FISH) detects enrichment of fusobacteria in colorectal tumors……75

Figure 2-4. Phylogenetic analysis identifies several *Fusobacterium* species in human colon cancer tissues…..77

Figure 3-1. *Fusobacterium nucleatum* promotes intestinal tumorigenesis and is enriched in tumor tissues of $Apc^{Min/+}$ mice……90

Figure 3-2. *F. nucleatum* selectively expands myeloid-derived immune cells, but not lymphoid immune cells in the intestinal tumor microenvironment……92

Figure 3-3. A *Fusobacterium*-associated human colorectal cancer gene signature shared and validated in mice……96

Figure 3-4. *Fusobacterium* is enriched in adenoma versus adjacent normal tissue and detected at a higher abundance in stool from CRC and adenoma cases than from healthy controls……...98

Supplementary Figure 1-1. Generation of artificial shotgun sequencing reads……118

Supplementary Figure 1-2. Applying a sequencing error distribution model to artificially generated reads.119

Supplementary Figure 1-3. Identification of virus-derived reads by sequence similarity……120

Supplementary Figure 1-4. Probability of contig formation for sequences randomly generated from a genome of varying size……121

Supplementary Figure 2-1. Phylum-level microbial classifications of whole-genome sequencing on nine colon tumor and normal tissue pairs……122

Supplementary Figure 2-2. Whole genome sequencing identifies the *Streptococcaceae* family as significantly enriched in colon tumors……123

Supplementary Figure 2-3. Phylum-level relative abundance of the microbiota by 16S rDNA sequencing on 95 colon tumor and normal tissue pairs……124

Supplementary Figure 2-4. Microbial diversity in the colon cancer microbiome……125

Supplementary Figure 2-5. *Fusobacterium* load correlates with geography……126

Supplementary Figure 2-6. Validation of the tumor-enrichment of *Fusobacterium* by quantitative PCR…..127

Supplementary Figure 2-7. Fluorescence *in situ* hybridization (FISH) detects enrichment of fusobacteria in colorectal tumors……128

Supplementary Figure 2-8. Quantitation of *Fusobacterium* across analytic methods……...…….129

Supplementary Figure 2-9. Relative abundance measurements for *Fusobacterium* OTUs in human colon cancer tissues…….130

Supplementary Figure 3-1. *F. nucleatum* does not affect intratumoral Th17 nor Foxp3+ Treg cell numbers.131

Supplementary Figure 3-2. Human colon tumor RNA-Seq analysis shows enrichment for inflammatory functions in genes correlated with *Fusobacterium* abundance but not the abundance of other genera……...132

List of Tables

Supplementary Table 1-1. Bacterial genomes used to construct an artificial sequence dataset to test the metagenomics module of PathSeq……...133

Supplementary Table 1-2. Metagenomic analysis on a sequence dataset constructed from a set of twelve bacterial genomes……..……134

Supplementary Table 1-3. PathSeq performance on artificially generated sequence data……..……135

Supplementary Table 1-4. PathSeq performance on artificially generated sequence data with introduced sequencing errors…….……136

Supplementary Table 2-1. PathSeq run-statistics and microbial classifications using colon cancer human whole-genome sequencing data……...……..137

Supplementary Table 2-2. Identification of viruses in colon cancer human whole-genome sequencing data by PathSeq analysis……138

Supplementary Table 2-3. Quantitative PCR for Fusobacterium in colon cancer cell lines……139

Supplementary Table 2-4. Quantitative PCR for *Fusobacterium* in metastases from colorectal carcinoma...140

Supplementary Table 2-5. Reference strain identification for Fusobacterium phylogenetic analysis……..141

Supplementary Table 3-1. Sample list of all colonic adenomas assessed for *Fusobacterium* abundance…….142

Supplementary Table 3-2. *Fusobacterium* is detected at a higher abundance in stool from CRC and adenoma cases than from healthy controls……143

List of Abbreviations

ACF	aberrant crypt foci
AOM	azoxymethane
***Apc*$^{Min/+}$**	multiple-intestinal neoplasia model of FAP bearing a heterozygous mutation in *APC*
BFT	*Bacteroides fragilis* toxin
CAC	colitis-associated cancer
CD	Crohn's disease
CFU	colony-forming unit
CLR	C-type lectin receptor
CRC	colorectal cancer
DC	dendritic cell
DCA	deoxycholic acid
DEN	diethylnitrosamine
DMBA	7,12-dimethylbenz[*a*]anthracene
DSS	dextran sulfate sodium
EDRN	Early Detection Research Network
ETBF	Enterotoxigenic *Bacteroides fragilis*
FAP	familial adenomatous polyposis
FISH	fluorescence *in situ* hybridization
GF	germ-free
HMP	Human Microbiome Project
HPV	human papillomavirus
IBD	inflammatory bowel disease
IEC	intestinal epithelial cell
LAB	lactic acid bacteria
LDA	Linear Discriminant Analysis
MALT	mucosa-associated lymphoid tissue
MAM	methylazoxymethanol
MCA	3′-methylcholanthrene
MetaHIT	Metagenomics of the Human Intestinal Tract
NLR	Nucleotide-binding domain leucine-rich repeat protein or NOD-like receptor
NSAID	non-steroidal anti-inflammatory drug
OTU	operational taxonomic unit
PGE2	prostaglandin E2
PRR	pattern recognition receptor
ROI	reactive oxygen intermediate
SCFA	short chain fatty acid
SFB	segmented filamentous bacteria
SMO	spermine oxidase
TCGA	The Cancer Genome Atlas

Th1	T helper type 1 T cell
Th17	T helper type 17 T cell
Th2	T helper type 2 T cell
TLR	Toll-like receptor
TPA	12-*O*-tetradecanoylphorbol 13-acetate
Treg	regulatory T cell
UC	ulcerative colitis
WT	wild-type

INTRODUCTION

Microbes and Inflammation in Cancer

Portions of this Introduction appear in the following publication:

Kostic, A.D., Howitt, M.R., Garrett, W.S. Exploring host-microbiota interactions in animal models and humans. Genes & Development (in press).

Overview

At the most fundamental level, a tumor is composed of a group of neoplastic cells harboring aberrant genomes that endow it with the ability to proliferate beyond normal means. Hanahan and Weinberg famously described the six hallmarks of cancer that most, if not all, tumors must acquire: self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis [1]. The focus of cancer research in the decades leading up to this seminal review had been on the cancer cell. But the tumor microenvironment, composed of not only the neoplastic cells themselves but also the surrounding stromal cells and a vast array of immune cells, may be just as important to the tumorigenic process as the cancer cell itself [2]. Malignant cells do not act alone in driving cancer, but rather they influence their surrounding cells to participate in the growth of the tumor. The past decade has seen tremendous advances in our understanding of the contribution of the microenvironment to tumorigenesis, and therefore it may not be surprising that the hallmarks of cancer have been recently updated to include four additional members that emphasize the role of the microenvironment: tumor-promoting inflammation, avoidance of immune destruction, deregulating cellular energetics, and genome instability and mutation [3].

The past decade has also seen a reimagining of the extent of our physiological interaction with our resident microbiota. Recent research has revealed roles for the microbiome in host physiology ranging from angiogenesis [4,5] and skeletal biology [6,7] to lipid metabolism [8,9] and even behavior [10-13]. Likewise, there is an increasing interest in the impact of the microbiota on cancer [14]. The lumen of the human distal gut is one of the most densely populated ecosystems on our planet, and we also harbor several distinct microbiomes on the

other surfaces of our body such as the respiratory and urogenital tracts and the skin. The number of bacterial cells on our body outnumber our own cells by a factor of ten, and from a functional standpoint, the number of genes encoded in our microbiome outnumbers the genes in our genome by a factor of 100. Therefore, the tumor microenvironment, in addition to containing neoplastic cells, stromal cells, and immune cells, should also be considered in the context of an additional component: the tumor microbiota.

In this Introduction I explore the roles of the microbiota and microbiota-associated inflammation in cancer. I use the colon as a model ecosystem for these interactions, although the microbiota and inflammation can have influences on cancer types beyond intestinal cancers.

The Human Gastrointestinal Microbiota

Overview

The meta'omics revolution in both deep sequencing and big data analytics is fostering an explosion of interest in how the gut microbiome impacts physiology and propensity to disease. Also driving interest in this field is the relatively recent discovery of the causal role of *Helicobacter pylori* in duodenal ulcers and gastric cancers [15,16], the role of the gut microbiota in inflammatory bowel disease [17], a new understanding of the impact of antibiotic use on the microbiota and the emergence of opportunistic pathogens [18,19], and the therapeutic use of fecal transplantation. Although the first successful application of fecal transplantation to *Clostridium difficile*-associated disease was first reported more than 50 years ago, the astounding clinical efficacy of this approach has only recently been demonstrated in a controlled trial [20].

The microbiota is necessary for the maintenance of physiologic homeostasis, but in the context of specific microbial and/or host phenotypes, the microbiota can contribute to disease

pathogenesis. For instance, the type VI secretion system (T6SS) of *Helicobacter hepaticus* directs an anti-inflammatory gene expression profile on intestinal epithelial cells (IECs) to create a tolerogenic immune environment, but *H. hepaticus* T6SS mutants disrupt this balance and promote colitis by driving a T helper type 17 T cell (Th17) response [21]. *H. hepaticus* is an example of what has been termed a pathobiont, a symbiotic microbe that is capable of promoting pathology only when specific host, environmental, or microbial factors are altered [22,23]. Similarly, *Helicobacter pylori* is a human pathobiont that is a resident of the stomach in 50% of the population globally [24] and may be protective against asthma and other allergic disease [18,25,26], but under specific conditions which have yet to be completely defined, *H. pylori* causes gastric cancer [16]. Host genetics also play a role in making a pathobiont pathogenic. For example, germ-free *Il10*$^{-/-}$ mice do not show any signs of intestinal inflammation, but when they are colonized by commensal strains of *E. coli* and *Enterococcus faecalis* that do not cause any disease in wild-type mice, these bacteria drive severe colitis in the IL-10-deficient mouse [27]. The emergence of pathobionts can also be influenced by diet [28] and by the use of antibiotics [29].

The Intestinal Ecosystem: The Flora

There is significant variability in the density and complexity of the microbiota throughout the body. Within the gastrointestinal tract, we harbor $10^{8\text{-}10}$ colony-forming units (CFU) of bacteria per gram (g) of saliva, 10^{3} CFU/g of gastric juice, $10^{2\text{-}4}$ CFU/g of contents in the duodenum and jejunum of the small intestine, 10^{10} CFU/g in the small intestinal ileum, and $10^{10\text{-}14}$ CFU/g of colonic content [14]. Interestingly, the higher microbial density in the colon relative to the small intestine correlates with a greater than 12-fold risk of cancer in the colon [30]. The

human gastrointestinal tract is home to more than 100 trillion bacteria, and the microbiome contains as many as 150 times the number of genes in the human genome [31,32].

The human intestinal microbiota is comprised predominantly of two phyla, the Firmicutes and Bacteroidetes, with a smaller representation of the Proteobacteria and Actinobacteria, and a rare representation of the Fusobacteria, Verrucomicrobia, and Cyanobacteria [30,33,34]. The mouse serves as an excellent model for microbiota studies because its gut microbiota bears striking similarities to ours [35]. The intestinal microbiota is dominated by strict anaerobes, most notably the genera of *Bacteroides*, *Eubacterium*, *Bifidobacterium*, *Peptostreptococcus*, and *Atopobium*, whereas facultative anaerobes are present at 1,000-fold lower levels [14]. Approximately 500 to 1,000 different species comprise the normal colonic microbiota [36].

Comprehensive characterizations of the "healthy" human adult microbiota have been carried out by the European and Chinese-led Metagenomics of the Human Intestinal Tract (MetaHIT) Consortium and the NIH-sponsored Human Microbiome Project (HMP). MetaHIT focused on shotgun metagenomic sequencing of fecal samples from 146 European individuals, uncovering the presence of a minimal gut genome and metagenome, based on taxonomy and gene functionality [37]. MetaHIT reported that individuals, regardless of gender, race, or geography, can be grouped into one of three enterotypes characterized by variation in the level of the *Bacteroides*, *Prevotella*, and *Ruminococcus* genera [38], however recent studies have favored a continuum or gradient of species rather than discrete enterotypes [39-41]. The HMP has, in a series of 16 articles published in parallel (www.plos-collections.org/hmp), produced the largest dataset to-date of the diversity of the human microbiome across body sites on a large number of healthy adults. The HMP carried out both 16S rRNA gene sequencing and metagenomic sequencing on 15 (for males) or 18 (for females) body sites on each of 242 healthy adults in the

United States sampled on three separate visits [42]. The gut and tooth habitats were found to harbor the greatest between-subject microbial diversity but also the lowest between-visit variability, whereas the skin had lower between-subject diversity but much higher between-visit variability [43]. These studies suggest that the composition of the microbiota varies substantially across individuals, though it is temporally stable within a single person, and the functionality of the microbiota at the gene-level is highly conserved across individuals [43].

The Intestinal Ecosystem: The Host

Though host-microbiota interactions are bidirectional, direct contact with the epithelium is limited by a thick mucus layer and secreted factors. The mucus layer is principally composed of mucin, a high molecular weight glycosylated protein, and also consists of trefoil peptides, antimicrobial peptides, and secretory IgA [44]. The colon has a 150μm inner mucus layer that is firmly adherent to the epithelium and a variable-thickness outer layer that is loose and non-attached, whereas the small intestine has a single-layer incomplete barrier [45,46]. α-defensins, small antibacterial peptides secreted by Paneth cells of the small intestine, are required for the maintenance of the composition of the luminal microbiota [47]. By contrast, RegIIIγ is a secreted antimicrobial lectin that does not affect the luminal microbiota, rather it is limited to the mucus layer and restricts the access of Gram-positive bacteria to the epithelium [48,49].

Colorectal Cancer

Gastrointestinal cancers account for 25% of all cancer incidences and for 9% of cancer deaths globally [36], and there are nearly 1 million new cases of colorectal cancer (CRC) worldwide each year (http://www.who.int/en). Familial CRC accounts for 15-20% of CRC cases [50], which are classified as familial adenomatous polyposis (FAP), MYH-associated polyposis,

hereditary nonpolyposis CRC, hamartomatous polyposis syndromes, or hyperplastic polyposis syndrome [51]. The genes responsible for the majority of these disorders have been uncovered: *MHL1*, *MSH2*, *MSH6*, *PMS2*, *APC*, *MYH*, *STK11*, *SMAD4*, *BMPR1A*, and *PTEN* [51]. One of the most widely used mouse models of CRC, the $Apc^{Min/+}$ mouse, bears a heterozygous mutation in the tumor suppressor gene *APC* [52,53], as do patients with FAP [54-56]. The $Apc^{Min/+}$ mouse will develop tens to hundreds of tumors in the small and large bowel [52], whereas FAP patients can develop thousands of adenomas but they are limited to the large bowel [57]. The identification of the germline mutations responsible for familial CRC has been instrumental in understanding the molecular pathways that underlie sporadic (i.e. non-familial) CRC.

CRC tumorigenesis proceeds through a series of genetic alterations, as first proposed by Fearon and Vogelstein [58]. The Wnt—β-catenin signaling pathway is essential for intestinal epithelial cell renewal, and mutations that lead to Wnt—β-catenin activation occur early in the course of tumorigenesis in greater than 90% of sporadic CRC cases; these mutations most commonly occur in *APC* and also *GSK3B*, which encodes a kinase that controls APC and β-catenin stability [59,60]. The hyperpoliferative epithelium will then accumulate additional mutations, most commonly in *KRAS* and *TP53*, as it increases in size, dysplasia, and villous content to form a malignant tumor [58]. Recent large scale cancer genome sequencing efforts have confirmed that *APC*, *TP53*, and *KRAS* are the most commonly mutated genes in non-hypermutated CRC tumors, but mutations in *PIK3CA*, *SMAD4*, and the F-box protein *FBXW7* are also frequently observed [61], as well as a recently-discovered recurrent *VTI1A-TCF7L2* fusion [62].

Colitis-associated cancer (CAC) is a form of colon cancer that is preceded by clinically detectable inflammatory bowel disease (IBD). IBD is classified into either ulcerative colitis

(UC), characterized by a non-transmural mucosal inflammation that is limited to the colon, and Crohn's disease (CD), which is a transmural inflammation of the mucosa that can affect the entire gastrointestinal tract [63,64]. Patients with CD have an 8% increased cumulative risk of CAC after 30 years of active CD, and for UC patients it is 18-20% [65]. The risk of developing CRC is 10-fold higher for individuals with IBD compared to the healthy population [30], however less than 2% of all CRC cases have a history of IBD [66]. Interestingly, wild-type (WT) mice can reproducibly develop CAC by administering the detergent dextran sulfate sodium (DSS) in the drinking water along with a single injection of the carcinogen azoxymethane (AOM) [67]. In contrast to sporadic CRC, mouse models of CAC develop mutations affecting Wnt—β-catenin signaling relatively late in the tumorigenic process, and develop early mutations in *Trp53* and *Kras* [60,68].

The Role of Inflammation in Colorectal Cancer

Inflammation is a hallmark of cancer [3]. Chronic inflammation can drive sustained innate immune cell recruitment, tumor growth, and metastasis, and directly promotes malignant cell transformation by inducing chromosomal and microsatellite instability, CpG island methylation, epigenetic alteration, and post-translational modifications [44]. Therefore, inflammation can contribute to all three stages of tumorigenesis: initiation, promotion, and progression [69].

Inflammation contributes to tumor initiation by virtue of its ability to cause mutations. Indeed, mucosal inflammation elicits systemic DNA damage that can contribute to tumoral genetic instability [70,71]. For example, mice deficient in ATM, a kinase involved in DNA double-strand break recognition and repair, show not only elevated DNA damage, but persistent

immune activation and increased sensitivity to DSS-induced colitis [72]. Another integral component of the DNA damage repair machinery, p53, is one of the most commonly mutated genes in CAC, but is also found to be mutated in colitic tissue without signs of dysplasia [73]. Inflammatory cells, particularly activated neutrophils and macrophages, produce significant amounts of reactive oxygen species and reactive nitrogen species that cause oxidative damage and can directly lead to oncogenic mutations in inflamed tissue [60].

Once a tumor has been initiated, there are numerous mechanisms by which inflammation can contribute to tumor promotion (i.e. proliferation) and progression (i.e. malignant transformation, invasion, and metastasis).

The Role of Inflammatory Responses to the Microbiota in Cancer

Pattern recognition receptors (PRRs) are the fingers of the immune surveillance system. They are an elegant set of receptors that can identify microbial ligands such as cell wall components or nucleic acids and are expressed on intestinal epithelial cells and mucosal immune cells. PRRs include Toll-like receptors (TLRs), Nucleotide-binding domain leucine-rich repeat proteins or NOD-like receptors (NLRs), and C-type lectin receptors (CLRs) among others [74-76]. TLRs and NLRs recognize the symbiotic microbiota, induce host defense responses against pathogens, and control adaptive immune responses. TLRs are transmembrane proteins containing leucine-rich repeats that play a crucial role in the innate immune response by sensing microbe-associated molecular patterns on bacteria, viruses, or parasites in the extracellular environment (TLRs 1, 2, 4-6, 11) or in endolysosomes (TLRs 3, 7-9, 10).

MyD88

MyD88 is an adaptor protein common to IL-1 and IL-18 signaling and to the TLRs (with the exception of TLR3). Thus, loss of MyD88 may be expected to impact a wide range of innate

immune sensing of the microbiota. The absence of MyD88 in the non-obese diabetic mouse strain led to an altered microbiota composition with enriched abundance of *Lactobacillaceae*, *Rikenellaceae*, and *Porphyromonadaceae* [77]. In another recent study, loss of MyD88 was examined from both a microbial ecology and host transcriptome perspective along the length of the small intestine and colon [78]. The small intestinal microbiota of MyD88-deficient mice was notable for an enrichment of segmented filamentous bacteria (SFB) and greater interindividual variation. Recently, SFB have garnered increased attention, as they promote a population of T cells [79], called T helper 17 cells, which function in immunity against extracellular bacteria and fungi. Targeted intestinal epithelial deletion of *MyD88* using Villin-Cre X *MyD88*-Flox mice has revealed that such mice have reduced levels of the polymeric immunoglobulin receptor mucin-2 and antibacterial peptides [80]. MyD88 also regulates of the expression of RegIIIγ, the antibacterial lectin that restricts the localization of bacteria and ensures its proper segregation from the inner mucus layer of the intestinal mucosa [48]. MyD88 has been demonstrated to have roles in promoting cancer as well. $MyD88^{-/-}$ mice formed fewer skin papillomas when exposed to the carcinogens 7,12-dimethylbenz[*a*]anthracene (DMBA) and 12-*O*-tetradecanoylphorbol 13-acetate (TPA), fewer fibrosarcomas when exposed to 3′-methylcholanthrene (MCA) [81], and fewer hepatic tumors in the diethylnitrosamine (DEN)-induced model of liver cancer [82] compared to WT mice. MyD88-deficient mice develop less inflammation and decreased colonic tumorigenesis compared to WT when treated with oxazalone-AOM [83] or DSS-AOM [84]. MyD88 is also instrumental in driving intestinal tumorigenesis in $Apc^{Min/+}$ mice. $Apc^{Min/+}$X $MyD88^{-/-}$ mice have significantly reduced small intestinal and colonic tumors relative to $Apc^{Min/+}$ mice [85] because MyD88 signaling post-transcriptionally stabilizes the c-myc protein through activation of the kinase ERK, which induces the multiple intestinal neoplasia phenotype [86].

TLRs

There have been several studies on the impact of the TLRs upstream of MyD88. Investigations of TLR5-deficient mice have revealed phenotypes resulting in metabolic syndrome [87]and colitis [88], with coincident alterations in the microbiota and transient elevations in the Proteobacteria and, in particular, enterobacterial species in colitic *Tlr5* knockout mice [87,89]. While there have been other observations supporting that alterations in TLR signaling impact the microbiota (e.g., that there is an altered colonic mucosal microbiota in *Tlr2*$^{-/-}$ mice) [90], other studies have challenged the magnitude of the impact of TLR signaling perturbations on the gut microbiome. A denaturing gradient gel electrophoresis and fluorescence *in situ* hybridization (FISH)-based investigation [91] and a more recent study using deep 16S rRNA gene surveys both call into question whether TLRs or MyD88 alter the gut microbiota in a genotype-dependent fashion [92]. Ubeda et al. (2012) generated MyD88-, TLR2-, TLR4-, TLR6-, and TLR9-deficient mouse lines from heterozygote X heterozygote breeding strategies. Interestingly, they did not detect statistically significant differences in community composition or diversity in cecal luminal, ileal luminal, or ileal microbial communities in these mice compared with their littermate controls. They also did not detect statistically significant differences in microbial community response after an antibiotic (vancomycin) perturbation. However, differences were observed between WT and TLR-mutant colonies that had been maintained as separate lines from homozygous X homozygous crosses for many years. These results raise awareness about the importance of considering lineage or legacy effects in microbiome studies, and it should be noted that these effects were considered in several of the cited studies [48]. Like its roles in shaping the microbiota, the impact of TLRs on intestinal tumorigenesis is also currently not well understood. Deletion of *Tlr4*, which is responsible for

detecting lipopolysaccharide of Gram-negative bacteria, protects DSS-AOM-treated mice from developing colon tumors [93,94], but deficiency of TLR2, which recognizes peptidoglycan and lipoteichoic acid among other molecules produced by bacteria and fungi, results in an increased intestinal tumor load [95].

NLRs and the Inflammasome

The other major class of PRRs are the NLRs. The NLRs are a central component of the inflammasome, which drives an innate immune response against intracellular pathogens [96,97]. The inflammasome is a multiprotein complex that is composed of several members of the NLR family, Procaspase-1, and the adaptor protein ASC [98]. Inflammasome activation results in Caspase-1 activation and subsequent Caspase-1 proteolytic activation of two proinflammatory cytokines: IL-1β and IL-18 [99]. Mutations in the NLR family member NOD2 increase the risk of Crohn's disease. Nod2 expression in the intestine is dependent on the presence of a gut microbiota, and *Nod2* knockout mice are more susceptible to colonization by intestinal mouse pathogens [100]. Mice deficient in Nod1 and Nod2 have altered gut microbiota composition as compared with their heterozygous littermates [101]. Nod1 is expressed ubiquitously in intestinal epithelial and immune cells, and recognizes the peptidoglycan of Gram-negative bacteria [102,103]. Nod1 has a protective role against colitis-associated cancer; DSS-AOM-treated $Nod1^{-/-}$ mice develop more colonic tumors than DSS-AOM-treated WT mice, and $Apc^{Min/+}$X $Nod1^{-/-}$ mice exhibit increased intestinal tumors than $Apc^{Min/+}$ controls [104]. NLRP6 is a NLR family member that functions in inflammasome, type 1 interferon, and NF-κB signaling. NLRP6 inflammasome-deficient mice have an altered gut microbiota notable for an expansion of *Prevotellaceae* and the TM7 phylum [105]. Although why this bacterial family expands remains unclear, this microbiota is functionally significant because co-housing and cross-fostering

experiments have revealed that the microbiota and the phenotype are transferable to WT mice, leading to increased colitis severity induced by DSS [105,106]. NLRP6 is a negative regulator of inflammatory signaling, and its loss activates MAP kinase and NF-κB signaling pathways downstream from TLRs and increases circulating monocyte numbers [107]. These cellular and signaling alterations may explain why NLRP6-deficient mice have an increased resistance to a number of pathogenic bacteria and an altered endogenous microbiota [107]. NLRP6 also has a role in suppressing inflammation-induced colon tumorigenesis, because *Nlrp6-/-* mice have increased colitis and colonic tumors when treated with DSS-AOM compared to WT controls [108,109]. Similar phenotypes are seen in mice deficient in the related NLR family members *Nlrp3* [110,111] and *Nlrp12* [112,113]. Caspases are cysteine proteases that play wide-ranging roles functioning in apoptosis to inflammation. Caspase-1 cleavage of IL-1β and IL-18 contributes to regulating inflammatory tone in the gut and thus may modulate the microbiome. The gut microbiota of Capsase-1-, Capsase-3-, and Capsase-7-deficient mice were recently evaluated in comparison with WT mice using 16S rRNA gene fecal profiling, and significant differences were observed across several families, including *Lachnospiraceae*, *Porphyromonodaceae*, and *Prevotellaceae* [114]. In keeping with the pivotal role of caspases in inflammasome function and consequently suppression of tumorigenesis, both Caspase-1-deficient [115] and Caspase-12-deficient [116] mice have increased susceptibility to developing colon tumors following treatment with DSS-AOM.

An intriguing insight that is gleaned from these studies is that, while both pathways are similarly involved in microbial detection and innate immune response, TLR/MyD88 signaling promotes the development of CRC, whereas NLR and inflammasome signaling appears to protect the host from CRC. This difference might be dictated by the motif that the PRR

recognizes or the location (i.e. extra- versus intracellular) of the PRR [30].

Defensins

TLRs and NLRs are of course not the only molecules involved in host response to the microbiota that affects tumorigenesis. Defensins are cationic proteins found in both the animal and plant kingdom that have broad antimicrobial activity against bacteria, fungi, and viruses, and the microbiota of mice with altered enteric defensin activity have been investigated. In mice and humans, defensins are principally produced by small intestinal Paneth cells and also by intestinal absorptive enterocytes [117]. Intestinal tissue and luminal samples from transgenic mice expressing one or two copies of the human α-defensin (DEFA5) and knockout mice deficient in the metalloprotease MMP7, which proteolytically activates α-defensin, were analyzed showed several significant differences in the gut microbiota from the phylum to the species level, including a strong reduction in the SFB [118]. Interestingly, α-defensin shows killing activity against *Helicobacter pylori* at low concentrations *in* vitro [119]. While it may seem intuitive that antimicrobial molecules, TLRs, and NLRs would impact the microbiome, the microbiome field is relatively young. Thus, there is a need to define the impact of such genes on the microbiota and clarify whether changes are stochastic or host genotype driven.

NF-κB Activation and the IL-6—STAT3 Axis

Inflammatory processes lead to the production of cytokines and growth factors that prevent malignant cells from apoptosis [120,121]. Genetic ablation of components of the NF-κB pathway or factors in STAT3 activation in epithelial cells blocks the expression of anti-apoptotic genes including Bcl-xL and Bcl-2, resulting in increased levels of apoptosis and more severe colitis, but remarkably, decreased tumor load [122-124]. When NF-κB activity was specifically ablated in myeloid cells with the use of IKKβ floxed mice and the LysMcre conditional deleter,

both colon tumor size and multiplicity were significantly reduced [122]. NF-κB activation in myeloid cells controls the expression of multiple inflammatory and tumor-promoting cytokines, including TNF-α and IL-1β [125,126], which are capable of activating NF-κB in epithelial and malignant cells [60].

NF-κB activation in myeloid cells also leads to the production of IL-6. IL-6 is a key cytokine in CAC; IL-6 enhances the proliferation of colonic carcinoma cells *in vitro*, is required for the survival of IECs and development of CAC *in vivo*, and interference with IL-6 signaling in the late stages of CAC results in the slowing of tumor growth [124,127,128]. IL-6 can influence the differentiation of Th17 cells (and IL-17 produced by Th17 cells in turn induce the IL-6—STAT3 signaling pathway [129], forming a positive feedback loop), the suppression of regulatory T cells (Tregs), regulate the recruitment of many myeloid cell subtypes, among other functions that influence the immune response [130,131].

IL-6 production by myeloid cells in turn drives STAT3 activation in epithelial cells, which can then also result in the activation of the Ras—Erk and PI3K—Akt pathways [60]. STAT3 leads to the up-regulation of anti-apoptotic genes (Bcl-xL and Bcl-2), cell cycle regulators (Cyclin D1, c-myc), and angiogenic factors (bFGF, VEGF) [121,132]. Conversely, the epithelial inactivation of STAT3 reduces cell survival and proliferation, and results in decreased CAC tumor growth and multiplicity [123,124].

IL-23

IL-23 is a member of the IL-12 cytokine family. IL-23 expression is up-regulated in many cancers including CRC [133,134], and IL-23 receptor (IL-23R) blockade reduces intestinal inflammation and tumor growth in $Apc^{Min/+}$ mice colonized with enterotoxigenic *Bacteroides fragilis* [135] (see *Associations between Single Bacterial Species and Cancer*). IL-23 may be a

key factor in driving intestinal tumorigenesis that results from epithelial barrier defects and infiltration of luminal bacteria into dysplastic regions [136] (see *Intestinal Barrier Defects and Microbial Infiltration in CRC*). IL-23 functions in part by up-regulating the Th17 response, and IL-23R blockade reduces IL-17A production [135], therefore it is likely that IL-23 imparts its effects by controlling the expression of IL-6, IL-22, and IL-17, but does not act on cancer cells directly [60].

COX-2

Cyclooxygenase-2 is an enzyme that converts arachidonic acid to prostaglandins, key mediators of inflammation [137]. Unlike COX-1, which is a housekeeping gene involved in producing prostaglandins at basal levels, COX-2 is not normally expressed by most cell types but can be strongly induced by a variety of different growth factors and pro-inflammatory cytokines [138]. Approximately 85% of human colorectal carcinomas and 50% of colonic adenomas exhibited elevated COX-2 expression [139-141], and COX-2 is up-regulated in intestinal adenomas from $Apc^{Min/+}$ mice [142]. Correspondingly, a daily dose of aspirin or other non-steroidal anti-inflammatory drug, which block COX-2 activity, over the course of 10-15 years can reduce the relative risk of developing CRC by up to 50% [143-146] and can reduce colonic adenoma size and number in FAP patients [147,148].

Mouse Models of Immunodeficiency in Colitis and Cancer

Several genetic mouse models of colitis have been interrogated regarding their microbiome patterns before, after, and during active inflammation, which can perturb the microbiota as oxidative stress and antimicrobial molecules increase during inflammation [149]. Mouse models of colitis, including *FVB.mdr1a*$^{-/-}$ [150], BALB/c.*T-bet*$^{-/-}$ X *Rag2*$^{-/-}$ [151,152], and *IL-10R2*$^{-/-}$ X *TgfbrII*$^{-/-}$ [153] mice, have been profiled using multiple methodologies inclusive of

culture-dependent and -independent techniques and revealed distinctive patterns with the gut microbiota. The *Enterobacteriaceae Klebsiella pneumoniae* and *Proteus mirabilis* were associated with colitis in *T-bet*$^{-/-}$ X *Rag2*$^{-/-}$ [152], while *Bacteroides thetaiotaomicron* was associated with disease in *IL-10R2*$^{-/-}$ X *TgfbrII*$^{-/-}$ mice [153], suggesting that a variety of genes may impact colonization and fitness of specific gut microbes in the setting of an inflamed epithelium and genetic immune perturbations. Remarkably, the colitigenic microbiota of the *T-bet*$^{-/-}$ X *Rag2*$^{-/-}$ mice is transmissible to WT mice by co-housing [151,152,154]. The *T-bet*$^{-/-}$ X *Rag2*$^{-/-}$ mice develop colonic adenocarcinoma after 6 months of age, but all signs of dysplasia are prevented by keeping the mice on a cocktail of antibiotics [155], suggesting that the microbiota is responsible for driving colonic tumorigenesis in this model.

The Role of the Gut Microbiota in Colorectal Cancer

Associations between Single Bacterial Species and Cancer

The best characterized association between a single bacterial species and cancer is that of *Helicobacter pylori* and gastric cancer, a discovery that resulted in a Nobel Prize for the co-discoverers Barry J. Marshall and J. Robin Warren [156]. *H. pylori* is the most common etiologic agent in infection-related cancer, and accounts for 5.5% of all cancers globally [157]. Most individuals that harbor *H. pylori*, approximately 50% of the world's population [24], do not develop peptic ulcers or cancer, but there is a significant amount of data to support a causal relationship between *H. pylori* and gastric cancer based largely on epidemiology and case-control studies [158]. A series of meta-analyses that assessed the association between gastric cancer and *H. pylori*, despite finding that approximately 50% of studies produced negative results, independently came to the conclusion that the association has an odds-ratio of

approximately 2.0 (range 1.92 to 2.56) [158-164]. Although there is conflicting evidence on whether or not *Helicobacter* eradication therapy reduces incidence of gastric cancer [165], at least one study convincingly demonstrates that successful eradication of *H. pylori* in gastric mucosa-associated lymphatic tissue lymphoma can successfully treat 80% of patients with early stage lymphomas [166]. Beyond epidemiological data, there is evidence for molecular mechanisms that link *H. pylori* infection with cancer. The development of gastric cancer is believed to be a multi-step process which includes superficial gastritis, chronic atrophic gastritis, and proceeds to metaplasia, dysplasia, and then carcinoma [167]. *H. pylori* is involved in the early stages of this process, causing chronic active gastritis and atrophic gastritis (Correa's hypothesis) [15,16]. Upon bacterial attachment to the epithelial cell, a number of bacterial proteins are secreted into the cell by the type IV secretion system, including CagA and the VacA toxin. VacA induces vacuole formation in the cell that stimulates apoptosis of the epithelial cell [168]. CagA becomes tyrosine-phosphorylated by endogenous kinases and then causes the sustained activation of SHP-2, ERK1/2, and Src kinase [169,170], leading to actin cytoskeletal changes in the cell that may promote proliferation [16].

There is a long-standing clinical observation that links *Streptococcus bovis* (now known as *Streptococcus gallolyticus*) infection with endocarditis and CRC [171]. Patients that present with *S. gallolyticus* endocarditis or septicemia are routinely screened for the presence of colorectal cancer, because 60% of *S. gallolyticus*-infected patients are found to have a concomitant adenoma or carcinoma [172]. The underlying pathophysiology of this strong association is not understood [173-175], but it may be the result of decreased epithelial barrier function at the site or colonic adenomas or carcinomas that allows streptococcal species to enter into the circulation [136].

Growing evidence suggests a role for adherent-invasive strains of *Escherichia coli* in driving CAC. Strains of *E. coli* that carry the polyketide synthase (*pks*) pathogenicity island cause DNA double-strand breaks and activation of the DNA damage checkpoint pathway, leading to cell cycle arrest and cell death *in vitro* [176]. The *pks* island encodes the polyketide-polypeptide genotoxin Colibactin. *E. coli* harboring Colibactin induced phosphorylated H2AX foci in mouse enterocytes *in vivo*, as well as the appearance of micronuclei, aneuploidy, ring chromosomes, and anaphase bridges, suggesting that Colibactin induces breakage-fusion-bridge cycles and chromosomal instability [177]. A recent study showed that monocolonizing germ-free, AOM-treated, *Il10*$^{-/-}$ mice with a *pks*-containing strain of *E. coli* resulted in enhanced tumor multiplicity compared to *Enterococcus faecalis* –monocolonized control mice despite similarly high levels of colitis [178]. Remarkably, intestinal tumorigenesis could be abolished in this model by deletion of the *pks* island in the *E. coli* strain [178]. Although previous studies have observed an association between *E. coli* and CRC [179] and Crohn's disease [180], *pks*-positive *E. coli* were found to be enriched in both IBD and CRC cohorts [178], suggesting that this DNA-damaging bacterium may be a clinically relevant etiologic agent in human CRC.

Enterotoxigenic *Bacteroides fragilis* (ETBF) secretes the metalloprotease toxin *B. fragilis* toxin (BFT) and is associated with inflammatory diarrheal disease in young children [181]. A stool-based quantitative PCR-based study found that ETBF was found at a higher prevalence in CRC patients compared to healthy individuals in a Turkish cohort [182]. This finding raised the possibility that ETBF could have a role in promoting tumorigenesis in this population, as other studies have shown that ETBF induces c-myc expression and proliferation when co-cultured with a colon carcinoma cell line [183], triggers the cleavage of E-cadherin and enhances β-catenin signaling *in vivo* [184], and that it induces colitis in WT mice [185]. In a seminal study, Cynthia

Sears and colleagues showed that *Apc*$^{Min/+}$ mice colonized with ETBF developed severe colitis and a significantly increased colonic tumor load compared to control mice colonized with a non-BFT-containing strain of *B. fragilis* [135]. The colonic lamina propria of ETBF-colonized mice were enriched for Th17 and γδ T lymphocytes, and antibody-based blockade of IL-17 ameliorated colonic tumorigenesis in this model [135]. Therefore, inflammation as a result of response to the BFT toxin drives inflammation and carcinogenesis. In addition to this inflammatory mechanism, ETBF was found to induce spermine oxidase (SMO), a polyamine catabolic enzyme, leading to the generation of reactive oxygen intermediates, and DNA damage *in vitro*, and inhibition of SMO reduces ETBF-mediated inflammation in *Apc*$^{Min/+}$ mice [186]. It has been proposed that ETBF, perhaps like *pks*-containing *E. coli*, is an "alpha-bug;" capable of introducing DNA damage, promoting IEC proliferation, driving tumor-promoting inflammation, and perhaps influencing the microbiota as whole to drive colonic tumorigenesis [187].

Evidence for a Role of the Microbiota in CRC from Gnotobiotic Mice

Gnotobiotics is the science of well-controlled microbial environments within and for biological specimens, encompassing the generation and maintenance of both germ-free (GF) and defined microbial community animals [188]. While gnotobiotic mice are discussed here, there are gnotobiotic fish, flies, rats, pigs, and foals. Gnotobiotic techniques have been essential to mechanistically interrogate host–microbiota interactions in mice [189]. Rederivation of any combination of genetic mutant mice is possible via embryo transfer into GF pseudopregnant mice or aseptic harvesting of a gestational uterine package and transfer of the fetuses to a GF foster female. It is important to realize that although GF mice do not harbor live bacteria or archaea, they are not naive to microbial-associated molecular patterns, as they encounter them in their sterilized (by autoclave or irradiation) food, water, and bedding. Microarray-based

comparisons of host tissues from GF mice and their conventional counterparts have been a successful approach to understand the broad impact of the microbiome on physiology [78,190,191]. Microarrays of GF and conventionalized mice even have shown that the gut microbiota modulate host gene expression post-transcriptionally by altering expression of host microRNAs within the small and large intestine [192].

The first evidence that the microbiota may have a role in CRC came from gnotobiotic experiments in GF rats that showed a higher incidence of CRC in conventionally-raised compared to GF rats [193]. Since then, there have been a number of studies in genetically engineered mice predisposed to CRC showing a lower incidence of disease under GF conditions. For instance, T cell receptor beta-chain and p53 double-knockout (*Tcrβ*$^{-/-}$ X *p53*$^{-/-}$) mice showed no incidence of intestinal adenocarcinoma under GF conditions, whereas ileocecal and cecal adenocarcinomas were detected in 70% of the conventionalized group [194]. *Tgf-β1*$^{-/-}$ mice normally develop CRC, but are free from inflammation, hyperplasia, and carcinoma when reared germ-free, but intestinal lesions re-appear when the mice are exposed to *Helicobacter hepaticus* [195]. Similarly, *Rag2*$^{-/-}$ mice, which lack an adaptive immune system, are also free from inflammation and all signs of hyperplasia under GF conditions, but develop inflammation and intestinal carcinoma in the presence of *Helicobacter hepaticus* [196]. *Apc*$^{Min/+}$ mice show a reduction in both small intestinal and colonic tumors under GF versus conventional conditions, and the gut microbiota has been found to trigger the c-Jun—JNK and STAT3 signaling pathways to accelerate tumor growth in this mouse model [197].

There are a number of studies that explore the contribution of the gut microbiota to colitis and CRC in IL-10-deficient mice. IL-10 was initially identified to be produced by T helper type 2 (Th2) cells, B cells, and macrophages to inhibit Th1 cell functions, and was subsequently

found to be a potent suppressor of macrophage activation and generally a strong anti-inflammatory cytokine [198,199]. Consequently, *Il10*$^{-/-}$ mice develop chronic colitis [198] and are also predisposed to developing CRC. Germ-free *Il10*$^{-/-}$ mice developed significantly less colitis than conventionally raised *Il10*$^{-/-}$ mice [200]. *Il10*$^{-/-}$ mice mono-associated with *Enterococcus faecalis* developed colitis and colonic adenocarcinoma, but both GF mice and mice mono-associated with a number of control strains of bacteria did not develop any signs of carcinoma [200]. IL-10-deficient mice that are treated with the carcinogen AOM develop colitis and CRC under conventional conditions, but do not develop any intestinal inflammation or dysplasia under GF conditions [201]. Two related studies suggest that colonization by *Helicobacter hepaticus*, a common symbiont of the mouse gut microbiota, specifically is required to drive AOM-induced colon tumors in *Il10*$^{-/-}$ mice [202,203]. Similarly, mono-association with a Colibactin-positive, but not Colibactin-negative, strain of *E. coli* in AOM-treated, *Il10*$^{-/-}$ mice drives CAC [178] (see *Associations between Single Bacterial Species and Cancer*). These studies indicate that the microbiota is required for the development of colonic tumors in a number of different mouse models and so raise the question: What microbial factors are responsible for contributing to tumorigenesis?

Microbial Products that Contribute to Tumorigenesis

Most studies on bacterial contributions to cancer have been approached from the perspective of the inflammatory response to the microbe (see *The Role of Inflammatory Responses to the Microbiota in Cancer*), but recent research points to possible direct contributions of microbial metabolites to the carcinogenic process.

Early work in rats showed that conventionally raised but not GF animals develop CRC when given the plant glycoside carcinogen cyasin, however GF rats develop colon tumors when

given the downstream active metabolite of cyasin, methylazoxymethanol (MAM) [204]. The conversion of cyasin to its carcinogenic form MAM was found to be imparted by β-glucosidase, an enzyme that is encoded by the microbiota [204]. In subsequent decades, it has been found that the microbiota can convert latent carcinogens into bioactive compounds through the action of a number of bacterial enzymes including β-glucosidase, β-glucuronidase, 7-α-dehydroxylase, azobenzene reductase, and nitroreductase [30,205].

β-glucuronidase

Many compounds that are metabolized in the liver become conjugated to glucuronic acid and then secreted into the small intestine via the bile. Bacterial β-glucuronidase in the colon hydrolyzes these conjugates, releasing the parent compound. For example, one of the most commonly used colonic carcinogens, AOM, is hydrolyzed into the activated procarcinogenic metabolite MAM in the liver, conjugated with glucuronic acid, and then secreted through the bile into the small intestine [206]. When this compound becomes introduced to the colonic microbiota, β-glucuronidase activity converts it to its carcinogenic form by removing the glucuronide group. The activity of additional, yet-uncharacterized bacterial enzymes also metabolize MAM into methyldiazonium and a reactive methyl-carbonium ion [207]. The inhibition of β-glucuronidase in AOM-treated rats significantly reduced colon tumor load [208]. In another example, the colon cancer chemotherapeutic CPT-11 causes severe diarrheal disease when reactivated into the toxic form by β-glucuronidase, and β-glucuronidase inhibitors block removal of the glucuronide group and thereby prevent toxicity caused by this substance [209]. As might be expected, populations at high risk for CRC show high levels of β-glucuronidase activity in stool [210]. Rats fed a high-fat diet, associated with a high risk for CRC, show increased β-glucuronidase activity in cecal and colonic contents [211]. Conversely, diets high in

fiber and bran are associated with decreased risk of CRC and decreased β-glucuronidase activity in rats [212] and in humans [213].

7-α-dehydroxylase

7-α-dehydroxylase is a microbial enzyme that converts the bile acid cholate into deoxycholic acid (DCA). DCA has been shown to cause DNA damage and apoptosis in IECs [214]. High levels of fecal DCA are associated with increased CRC risk [215], and 7-α-dehydroxylase activity is higher in stool from CRC patients relative to healthy control individuals [216]. Bile acids can disrupt the integrity of the colonocyte cell membrane, resulting in the release of arachidonic acid and prostaglandin E2 (PGE2), which can drive proliferation in IECs. Secondary bile acids can also contribute to cell necrosis, hyperplasia, proliferation, DNA damage, and apoptosis [14].

Reactive Oxygen Intermediates

Reactive oxygen intermediates are derivatives of molecular oxygen and include superoxides, hydrogen peroxide, hypochlorous acid, singlet oxygen, and hydroxyl radicals. Some intestinal microbes, for example *Enterococcus faecalis*, produce substantial amounts of extracellular superoxide and hydrogen peroxide that have been shown to damage DNA and promote chromosomal instability in IECs *in vitro* [217-219]. The use of superoxide dismutase and γ-tocopherol, which block the production of ROI, in co-cultures of *E. faecalis* and IECs reduced levels of chromosomal instability [218]. Of course *in vivo*, activated macrophages and neutrophils [220-222] as well as epithelial cells and pre-malignant cells [223,224] also produce significant amounts of ROI, and therefore it can be difficult to determine how much ROI is contributed by the host versus the microbes themselves. During chronic inflammation, the DNA damage caused by ROI leads to mutations, deletions, and sister chromatid exchanges, which all

contribute to chromosomal instability and drive the carcinogenic process [14].

Approaches aimed at reducing the production of ROI or the inhibition of nitric oxide synthase, which produces reactive nitrogen species, have been successful at reducing DNA damage and preventing tumorigenesis [221,223-225]. Mice that are deficient in both glutathione peroxidase enzymes Gpx1 and Gpx2, the major enzymes responsible for reducing hydroperoxides in the intestine, develop cancer with a penetrance of 25% under conventional conditions, less than 9% under specific pathogen-free conditions, and no tumors were observed under GF conditions [226].

N-nitroso Compounds, Heterocyclic Amines, and other Products

Nitrate in the diet is converted to nitrite by the intestinal microbiota [14], and nitrite reacts with amines, amides, and methylurea to produce nitric oxide compounds, carcinogenic DNA-alkylating agents [227,228]. Furthermore, experiments using germ-free rat have demonstrated that *N*-nitrosation in the colon is dependent on the gut microflora, the products of which have been shown to have carcinogenic properties [229,230]. Heterocyclic amines that are produced when meat is cooked at high temperatures have been implicated in a number of gastrointestinal cancers [30]. Recent studies have shown that the production of the active carcinogenic form of these compounds is dependent upon the enzymatic activity of the intestinal microbiota [231-234].

Certain members of the microbiota that have been demonstrated to promote tumorigenesis produce carcinogenic compounds. Specifically, enterotoxigenic *B. fragilis* produces *B. fragilis* toxin, a metalloproteinase that cleaves E-cadherin [185] and contributes to colonic tumorigenesis in a mouse model [135], strains of *E. coli* that encode the *pks* pathogenicity island can induce DNA double strand breaks in eukaryotic cells and promote CRC

[176-178], and Helicobacter pylori encodes CagA and VacA among other secreted proteins that contribute to carcinogenesis in gastric adenocarcinoma [15,164,235-237] (see *Associations between Single Bacterial Species and Cancer*).

Bacteria that Show Protective Effects in CRC

Just as certain members of the gut microbiota promote carcinogenesis, other bacteria show evidence of having a protective effect against cancer. For example, the daily administration of *Lactobacillus acidophilus* to $Apc^{Min/+}$ mice results in decreased tumor multiplicity and size [238]. Similarly, AOM-treated rats fed *L. acidophilus* had decreased aberrant crypt foci (ACF) [239], *L. acidophilus* in combination with *Bifidobacterium longum* inhibited colonic tumorigenesis induced by dimethylhydrazine (a derivative of AOM) [240,241], and administration of *Lactobacillus rhamnosus* with *Bifidobacterium lactis* in AOM-treated rats decreased colonic tumor load [242]. Some bacteria can inhibit the activity of enzymes that produce carcinogens. For example, AOM-treated rats that were fed *B. longum* and an inulin derivative had decreased colonic ACF and reduced β-glucuronidase activity [243]. The co-administration of *L. acidophilus* and *Lactobacillus casei* can decrease the enzymatic activity of β-glucuronidase, azoreductase, and nitroreductase in rats [244] and in humans [245]. *L. casei* can prevent DNA damage and tumorigenesis induced by the carcinogen methylnitronitrosoguanidine [246], and can metabolize a number of heterocyclic amines [247], and thereby prevent DNA damage in the colon and liver of rats [248]. Similarly, *B. longum* has been shown to have protective effects in rats against heterocyclic amine-induced colon, mammary, and liver carcinogenesis [249]. The high-potency probiotic preparation CSL#3 reduces signs of colitis and delays dysplasia in a rat model of CAC [250,251], and administration of the probiotic combination of *L. rhamnosus* GG and *B. lactis* Bb12 improved clinical biomarkers in human

colon cancer patients [252]. In addition to the above *in vivo* studies, there are many *in vitro* studies demonstrating anti-proliferative effects of specific bacteria: *Bifidobacterium adolescentis* can inhibit the production of TNF-α and the activity of β -glucuronidase and β -glucosidase, and reduce the proliferation of several human cell lines [253]; *Bacillus polyfermenticus* adheres to Caco-2 cells and confers a dose-dependent inhibition of proliferation [254]; Strains of lactic acid bacteria (LAB) reduce the growth and viability of HT-29 cells [255]; A number of *Bifidobacterium* species and *Lactobacillus* species can inhibit the growth of the MCF7 breast cancer cell line [256].

Probiotics

Probiotics are living microbes that can contribute to the health of the host. The species of probiotics that are most frequently used are *Lactobacillus* spp., *Bifidobacterium* spp., and LAB [36] and, in fact, most of the bacteria that show evidence of a protective effect against cancer are commonly-used strains of probiotics. Generally, probiotics do not become established members of the microbiota, but rather only persist during the time of dosage and shortly thereafter [257]. There are several mechanisms by which probiotics might impart a protective effect against carcinogenesis: the binding or degradation of carcinogens; the inactivation of microbial enzymes involved in procarcinogen activation; the production of anti-tumorigenic compounds; direct inhibitory effects on proliferation; the enhancement of the anti-tumor immune response; the production of biofilms that prevent the attachment or invasion of pathogenic bacteria; out-competing pathogenic bacteria [36,258].

A principle metabolic product of probiotic bacteria, and the gut microbiota as a whole, is the short chain fatty acid (SCFA), which includes butyrate, propionate, and acetate. These are produced by the fermentation of dietary fiber and starch. SCFA serves as the primary energy

source for colonocytes [259], influences immune responses and protects against inflammatory disease [260], and is involved in recovery from intestinal injury [259,260]. SCFA promotes cell cycle arrest and apoptosis in cultured IECs [261] by the up-regulation of Bak and down-regulation of Bcl-xL through the mitochondrial pathway of apoptosis [262,263]. Butyrate induces histone hyperacetylation [264] by inhibiting histone deacetylase activity [265,266], thus explaining in part its pro-apoptotic effects [267]. Therefore, SFCA is likely a key molecule that contributes protective effects against cancer by probiotics.

There is a 15 to 20 year time interval from the appearance of pre-cancerous lesions to the development of carcinoma in colon cancer, so although much work remains to determine the full extent of its efficacy, probiotics may be an effective therapy to prevent or delay CRC in patients that are predisposed to the disease [36,268-270].

Intestinal Barrier Defects and Microbial Infiltration in CRC

A unique feature of dysplastic intestinal lesions is the local loss of intestinal barrier function. This is explained in part by the lack of mucus production because of decreased differentiation into goblet cells, and also by the loss of tight junctions between epithelial cells. As a result, dysplastic lesions are prone to infiltration by the luminal bacteria of the intestine. Indeed, disruption of the mucus layer renders mice susceptible to intestinal inflammation [45,271,272], which progresses to CAC [273] and the accelerated adenoma growth that is seen in the context of the $Apc^{Min/+}$ genotype [274]. Interestingly, the absence of local barrier function at tumor sites in $Apc^{Min/+}$ mice results in the translocation of microbial products across the epithelium and increased IL-23 expression [136]. IL-23 is a key cytokine in driving tumorigenesis in response to microbial infiltration, because its inactivation results in decreased tumor number and size and reduced levels of other pro-inflammatory cytokines [136].

Thesis Summary

This thesis reports on the characterization of the colorectal tumor microbiota and describes the functional contribution of members of the microbiota to the tumor microenvironment. In Chapter 1, we introduce PathSeq, a computational tool that is designed to analyze deep-sequencing datasets such as human whole-genomes and transcriptomes for the presence of microorganisms. PathSeq removes human sequences by computational subtraction before identifying microbe-derived sequences, and therefore is capable of identifying the presence of any microorganism including previously unsequenced, novel microorganisms. PathSeq also includes a module that performs a metagenomic analysis by calculating the relative abundance of each bacterial taxon present in the sample. In Chapter 2, we use the metagenomic analysis module of PathSeq to characterize the colorectal cancer microbiome using whole-genome sequencing of CRC tumor and adjacent-normal tissue pairs. As a result of this analysis, we discovered a strong enrichment of *Fusobacterium* species in CRC tumors. Finally, in Chapter 3, we introduce *Fusobacterium* species and other members of the CRC microbiome to $Apc^{Min/+}$ mice, and find that *Fusobacterium* induces an NF-κB inflammatory response in the tumor microenvironment and accelerates intestinal tumorigenesis.

References

1. Hanahan D, Weinberg RA: **The hallmarks of cancer.** *Cell* 1999, **100**:57–70.

2. Hanahan D, Coussens LM: **Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment**. *Cancer Cell* 2012, **21**:309–322.

3. Hanahan D, Weinberg RA: **Hallmarks of Cancer: The Next Generation**. *Cell* 2011, **144**:646–674.

4. Stappenbeck TS, Hooper LV, Gordon JI: **Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells.** *Proc. Natl. Acad. Sci. U.S.A.* 2002, **99**:15451–15455.

5. Reinhardt C, Bergentall M, Greiner TU, Schaffner F, Östergren-Lundén G, Petersen LC, Ruf W, Bäckhed F: **Tissue factor and PAR1 promote microbiota-induced intestinal vascular remodelling**. *Nature* 2012, doi:10.1038/nature10893.

6. Sjögren K, Engdahl C, Henning P, Lerner UH, Tremaroli V, Lagerquist MK, Bäckhed F, Ohlsson C: **The gut microbiota regulates bone mass in mice.** *J. Bone Miner. Res.* 2012, **27**:1357–1367.

7. Cho I, Yamanishi S, Cox L, Methé BA, Zavadil J, Li K, Gao Z, Mahana D, Raju K, Teitler I, et al.: **Antibiotics in early life alter the murine colonic microbiome and adiposity**. *Nature* 2012, **488**:621–626.

8. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung Y-M, et al.: **Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease.** *Nature* 2011, **472**:57–63.

9. Semova I, Carten JD, Stombaugh J, Mackey LC, Knight R, Farber SA, Rawls JF: **Microbiota Regulate Intestinal Absorption and Metabolism of Fatty Acids in the Zebrafish**. *Cell Host Microbe* 2012, **12**:277–288.

10. Sudo N, Chida Y, Aiba Y, Sonoda J, Oyama N, Yu X-N, Kubo C, Koga Y: **Postnatal microbial colonization programs the hypothalamic-pituitary-adrenal system for stress response in mice.** *J. Physiol. (Lond.)* 2004, **558**:263–275.

11. Diaz Heijtz R, Wang S, Anuar F, Qian Y, Björkholm B, Samuelsson A, Hibberd ML, Forssberg H, Pettersson S: **Normal gut microbiota modulates brain development and behavior.** *Proceedings of the National Academy of Sciences* 2011, **108**:3047–3052.

12. Bravo JA, Forsythe P, Chew MV, Escaravage E, Savignac HM, Dinan TG, Bienenstock J, Cryan JF: **Ingestion of Lactobacillus strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve**. *Proceedings of the National Academy of Sciences* 2011, doi:10.1073/pnas.1102999108.

13. Ezenwa VO, Gerardo NM, Inouye DW, Medina M, Xavier JB: **Animal Behavior and the Microbiome**. *Science* 2012, **338**:198–199.

14. Zhu Q, Gao R, Wu W, Qin H: **The role of gut microbiota in the pathogenesis of colorectal cancer**. *Tumor Biol.* 2013, doi:10.1007/s13277-013-0684-4.

15. Peek RM, Blaser MJ: **Helicobacter pylori and gastrointestinal tract adenocarcinomas.** *Nat. Rev. Cancer* 2002, **2**:28–37.

16. Wroblewski LE, Peek RM, Wilson KT: **Helicobacter pylori and gastric cancer: factors that modulate**

disease risk. *Clin. Microbiol. Rev.* 2010, **23**:713–739.

17. Manichanh C, Borruel N, Casellas F, Guarner F: **The gut microbiota in IBD**. *Nat Rev Gastroenterol Hepatol* 2012, **9**:599–608.

18. Blaser M: **Antibiotic overuse: Stop the killing of beneficial bacteria**. *Nature* 2011, **476**:393–394.

19. Shanahan F: **The colonic microbiota in health and disease.** *Curr. Opin. Gastroenterol.* 2013, **29**:49–54.

20. van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, Visser CE, Kuijper EJ, Bartelsman JFWM, Tijssen JGP, et al.: **Duodenal Infusion of Donor Feces for Recurrent Clostridium difficile**. *N. Engl. J. Med.* 2013, doi:doi: 10.1056/NEJMoa1205037.

21. Chow J, Mazmanian SK: **A pathobiont of the microbiota balances host colonization and intestinal inflammation.** *Cell Host Microbe* 2010, **7**:265–276.

22. Mazmanian SK, Round JL, Kasper DL: **A microbial symbiosis factor prevents intestinal inflammatory disease.** *Nature* 2008, **453**:620–625.

23. Round JL, Mazmanian SK: **The gut microbiota shapes intestinal immune responses during health and disease.** *Nat. Rev. Immunol.* 2009, **9**:313–323.

24. De Vries AC, Van Driel HF, Richardus JH, Ouwendijk M, Van Vuuren AJ, De Man RA, Kuipers EJ: **Migrant communities constitute a possible target population for primary prevention of Helicobacter pylori-related complications in low incidence countries.** *Scand. J. Gastroenterol.* 2008, **43**:403–409.

25. Blaser MJ, Chen Y, Reibman J: **Does Helicobacter pylori protect against asthma and allergy?** *Gut* 2008, **57**:561–567.

26. Arnold IC, Dehzad N, Reuter S, Martin H, Becher B, Taube C, Müller A: **Helicobacter pylori infection prevents allergic asthma in mouse models through the induction of regulatory T cells**. *J. Clin. Invest.* 2011, **121**:3088–3093.

27. Kim SC, Tonkonogy SL, Karrasch T, Jobin C, Sartor RB: **Dual-association of gnotobiotic IL-10-/- mice with 2 nonpathogenic commensal bacteria induces aggressive pancolitis.** *Inflamm. Bowel Dis.* 2007, **13**:1457–1466.

28. Devkota S, Wang Y, Musch MW, Leone V, Fehlner-Peach H, Nadimpalli A, Antonopoulos DA, Jabri B, Chang EB: **Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10−/− mice**. *Nature* 2012, doi:10.1038/nature11225.

29. Ayres JS, Trinidad NJ, Vance RE: **Lethal inflammasome activation by a multidrug-resistant pathobiont upon antibiotic disruption of the microbiota.** *Nat. Med.* 2012, **18**:799–806.

30. Arthur JC, Jobin C: **The struggle within: Microbial influences on colorectal cancer**. *Inflamm. Bowel Dis.* 2010, **17**:396–409.

31. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE: **Metagenomic analysis of the human distal gut microbiome.** *Science* 2006, **312**:1355–1359.

32. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR: **Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases.** *Proc. Natl. Acad. Sci. U.S.A.* 2007, **104**:13780–13785.

33. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA: **Diversity of the human intestinal microbial flora.** *Science* 2005, **308**:1635–1638.

34. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, et al.: **A core gut microbiome in obese and lean twins.** *Nature* 2008, **457**:480–484.

35. Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI: **Obesity alters gut microbial ecology.** *Proc. Natl. Acad. Sci. U.S.A.* 2005, **102**:11070–11075.

36. Orlando A, Russo F: **Intestinal Microbiota, Probiotics and Human Gastrointestinal Cancers**. *Tumor Biol.* 2012, doi:10.1007/s12029-012-9459-1.

37. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al.: **A human gut microbial gene catalogue established by metagenomic sequencing.** *Nature* 2010, **464**:59–65.

38. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, et al.: **Enterotypes of the human gut microbiome.** *Nature* 2011, **473**:174–180.

39. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, Harris HMB, Coakley M, Lakshminarayanan B, O'Sullivan O, et al.: **Gut microbiota composition correlates with diet and health in the elderly**. *Nature* 2012, **488**:178–184.

40. Jeffery IB, Claesson MJ, O'Toole PW, Shanahan F: **Categorization of the gut microbiota: enterotypes or gradients?** *Nature Reviews Microbiology* 2012, **10**:591–592.

41. Koren O, Knights D, Gonzalez A, Waldron L, Segata N, Knight R, Huttenhower C, Ley RE: **A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets.** *PLoS Comput. Biol.* 2013, **9**:e1002863.

42. Consortium THMP: **A framework for human microbiome research**. *Nature* 2012, **486**:215–221.

43. Consortium THMP: **Structure, function and diversity of the healthy human microbiome**. *Nature* 2012, **486**:207–214.

44. Cario E: **Microbiota and innate immunity in intestinal inflammation and neoplasia.** *Curr. Opin. Gastroenterol.* 2013, **29**:85–91.

45. Johansson MEV, Phillipson M, Petersson J, Velcich A, Holm L, Hansson GC: **The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria.** *Proceedings of the National Academy of Sciences* 2008, **105**:15064–15069.

46. Johansson MEV, Larsson JMH, Hansson GC: **The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions.** *Proceedings of the National Academy of Sciences* 2011, **108 Suppl 1**:4659–4665.

47. Salzman NH, Ghosh D, Huttner KM, Paterson Y, Bevins CL: **Protection against enteric salmonellosis in transgenic mice expressing a human intestinal defensin**. *Nature* 2003, **422**:522–526.

48. Vaishnava S, Yamamoto M, Severson KM, Ruhn KA, Yu X, Koren O, Ley R, Wakeland EK, Hooper LV: **The Antibacterial Lectin RegIII{gamma} Promotes the Spatial Segregation of Microbiota and Host in the Intestine [Internet]**. *Science* 2011, **334**:255–258.

49. Hooper LV, Littman DR, Macpherson AJ: **Interactions Between the Microbiota and the Immune System**. *Science* 2012, **336**:1268–1273.

50. Castells A, Castellví-Bel S, Balaguer F: **Concepts in familial colorectal cancer: where do we stand and what is the future?** *Gastroenterology* 2009, **137**:404–409.

51. Rustgi AK: **The genetics of hereditary colon cancer.** *Genes & Development* 2007, **21**:2525–2538.

52. Moser AR, Pitot HC, Dove WF: **A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse.** *Science* 1990, **247**:322–324.

53. Su LK, Kinzler KW, Vogelstein B, Preisinger AC, Moser AR, Luongo C, Gould KA, Dove WF: **Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene.** *Science* 1992, **256**:668–670.

54. Kinzler KW, Nilbert MC, Su LK, Vogelstein B, Bryan TM, Levy DB, Smith KJ, Preisinger AC, Hedge P, McKechnie D: **Identification of FAP locus genes from chromosome 5q21.** *Science* 1991, **253**:661–665.

55. Nishisho I, Nakamura Y, Miyoshi Y, Miki Y, Ando H, Horii A, Koyama K, Utsunomiya J, Baba S, Hedge P: **Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients.** *Science* 1991, **253**:665–669.

56. Groden J, Thliveris A, Samowitz W, Carlson M, Gelbert L, Albertsen H, Joslyn G, Stevens J, Spirio L, Robertson M: **Identification and characterization of the familial adenomatous polyposis coli gene.** *Cell* 1991, **66**:589–600.

57. Fearon ER: **Molecular genetics of colorectal cancer.** *Annu Rev Pathol* 2011, **6**:479–507.

58. Fearon ER, Vogelstein B: **A genetic model for colorectal tumorigenesis**. *Cell* 1990, **61**:759–767.

59. Schneikert J, Behrens J: **The canonical Wnt signalling pathway and its APC partner in colon cancer development.** *Gut* 2007, **56**:417–425.

60. Grivennikov S: **Inflammation and colorectal cancer: colitis-associated neoplasia**. *Semin Immunopathol* 2013, **35**:229–244–244.

61. Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, Kovar CL, Lewis LR, Morgan MB, Newsham IF, et al.: **Comprehensive molecular characterization of human colon and rectal cancer**. *Nature* 2012, **487**:330–337.

62. Bass AJ, Lawrence MS, Brace LE, Ramos AH, Drier Y, Cibulskis K, Sougnez C, Voet D, Saksena G, Sivachenko A, et al.: **Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion**. *Nat. Genet.* 2011, doi:10.1038/ng.936.

63. Baumgart DC, Sandborn WJ: **Inflammatory bowel disease: clinical aspects and established and evolving therapies.** *Lancet* 2007, **369**:1641–1657.

64. Xavier RJ, Podolsky DK: **Unravelling the pathogenesis of inflammatory bowel disease.** *Nature* 2007, **448**:427–434.

65. Rubin DC, Shaker A, Levin MS: **Chronic intestinal inflammation: inflammatory bowel disease and colitis-associated colon cancer.** *Front Immunol* 2012, **3**:107.

66. Mattar MC, Lough D, Pishvaian MJ, Charabaty A: **Current management of inflammatory bowel disease and colorectal cancer.** *Gastrointest Cancer Res* 2011, **4**:53–61.

67. Clapper ML, Cooper HS, Chang W-CL: **Dextran sulfate sodium-induced colitis-associated neoplasia: a promising model for the development of chemopreventive interventions**. *Acta Pharmacologica Sinica*

2007, **28**:1450–1459.

68. Terzić J, Grivennikov S, Karin E, Karin M: **Inflammation and Colon Cancer**. *Gastroenterology* 2010, **138**:2101–2114.e5.

69. Grivennikov SI, Greten FR, Karin M: **Immunity, Inflammation, and Cancer**. *Cell* 2010, **140**:883–899.

70. Westbrook AM, Wei B, Braun J, Schiestl RH: **Intestinal mucosal inflammation leads to systemic genotoxicity in mice.** *Cancer Res.* 2009, **69**:4827–4834.

71. Westbrook AM, Wei B, Braun J, Schiestl RH: **Intestinal inflammation induces genotoxicity to extraintestinal tissues and cell types in mice.** *Int J Cancer* 2011, doi:10.1002/ijc.26146.

72. Westbrook AM, Schiestl RH: **Atm-deficient mice exhibit increased sensitivity to dextran sulfate sodium-induced colitis characterized by elevated DNA damage and persistent immune activation.** *Cancer Res.* 2010, **70**:1875–1884.

73. Kraus S, Arber N: **Inflammation and colorectal cancer**. *Current Opinion in Pharmacology* 2009, **9**:405–410.

74. Geijtenbeek TBH, van Vliet SJ, Engering A, 't Hart BA, van Kooyk Y: **Self- and Nonself-Recognition by C-Type Lectins on Dendritic Cells**. *Annu. Rev. Immunol.* 2013, **22**:33–54.

75. Abreu MT, Fukata M, Arditi M: **TLR signaling in the gut in health and disease.** *J. Immunol.* 2005, **174**:4453–4460.

76. Rakoff-Nahoum S, Medzhitov R: **Innate immune recognition of the indigenous microbial flora.** *Mucosal Immunol* 2008, **1 Suppl 1**:S10–4.

77. Shen XJ, Rawls JF, Randall T, Burcal L, Mpande CN, Jenkins N, Jovov B, Abdo Z, Sandler RS, Keku TO: **Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas.** *Gut Microbes* 2010, **1**:138–147.

78. Larsson E, Tremaroli V, Lee YS, Koren O, Nookaew I, Fricker A, Nielsen J, Ley RE, Backhed F: **Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88**. *Gut* 2011, doi:10.1136/gutjnl-2011-301104.

79. Ivanov II, Atarashi K, Manel N, Brodie EL, Shima T, Karaoz U, Wei D, Goldfarb KC, Santee CA, Lynch SV, et al.: **Induction of Intestinal Th17 Cells by Segmented Filamentous Bacteria**. *Cell* 2009, **139**:485–498.

80. Frantz AL, Rogier EW, Weber CR, Shen L, Cohen DA, Fenton LA, Bruno MEC, Kaetzel CS: **Targeted deletion of MyD88 in intestinal epithelial cells results in compromised antibacterial immunity associated with downregulation of polymeric immunoglobulin receptor, mucin-2, and antibacterial peptides.** *Mucosal Immunol* 2012, **5**:501–512.

81. Swann JB, Vesely MD, Silva A, Sharkey J, Akira S, Schreiber RD, Smyth MJ: **Demonstration of inflammation-induced cancer and cancer immunoediting during primary tumorigenesis.** *Proceedings of the National Academy of Sciences* 2008, **105**:652–656.

82. Naugler WE, Sakurai T, Kim S, Maeda S, Kim K, Elsharkawy AM, Karin M: **Gender disparity in liver cancer due to sex differences in MyD88-dependent IL-6 production.** *Science* 2007, **317**:121–124.

83. Schiechl G, Bauer B, Fuss I, Lang SA, Moser C, Ruemmele P, Rose-John S, Neurath MF, Geissler EK, Schlitt H-J, et al.: **Tumor development in murine ulcerative colitis depends on MyD88 signaling of**

colonic F4/80+CD11b(high)Gr1(low) macrophages. *J. Clin. Invest.* 2011, **121**:1692–1708.

84. Salcedo R, Worschech A, Cardone M, Jones Y, Gyulai Z, Dai R-M, Wang E, Ma W, Haines D, O'hUigin C, et al.: **MyD88-mediated signaling prevents development of adenocarcinomas of the colon: role of interleukin 18.** *Journal of Experimental Medicine* 2010, **207**:1625–1636.

85. Rakoff-Nahoum S, Medzhitov R: **Regulation of spontaneous intestinal tumorigenesis through the adaptor protein MyD88.** *Science* 2007, **317**:124–127.

86. Lee SH, Hu L-L, Gonzalez-Navajas J, Seo GS, Shen C, Brick J, Herdman S, Varki N, Corr M, Lee J, et al.: **ERK activation drives intestinal tumorigenesis in Apcmin/+ mice**. *Nat. Med.* 2010, **16**:665–670.

87. Vijay-Kumar M, Aitken JD, Carvalho FA, Cullender TC, Mwangi S, Srinivasan S, Sitaraman SV, Knight R, Ley RE, Gewirtz AT: **Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5.** *Science* 2010, **328**:228–231.

88. Vijay-Kumar M, Sanders CJ, Taylor RT, Kumar A, Aitken JD, Sitaraman SV, Neish AS, Uematsu S, Akira S, Williams IR, et al.: **Deletion of TLR5 results in spontaneous colitis in mice.** *J. Clin. Invest.* 2007, **117**:3909–3921.

89. Carvalho FA, Koren O, Goodrich JK, Johansson MEV, Nalbantoglu I, Aitken JD, Su Y, Chassaing B, Walters WA, Gonzalez A, et al.: **Transient Inability to Manage Proteobacteria Promotes Chronic Gut Inflammation in TLR5-Deficient Mice**. *Cell Host Microbe* 2012, **12**:139–152.

90. Kellermayer R, Dowd SE, Harris RA, Balasa A, Schaible TD, Wolcott RD, Tatevian N, Szigeti R, Li Z, Versalovic J, et al.: **Colonic mucosal DNA methylation, immune response, and microbiome patterns in Toll-like receptor 2-knockout mice.** *FASEB J.* 2011, **25**:1449–1460.

91. Loh G, Brodziak F, Blaut M: **The Toll-like receptors TLR2 and TLR4 do not affect the intestinal microbiota composition in mice.** *Environ. Microbiol.* 2008, **10**:709–715.

92. Ubeda C, Lipuma L, Gobourne A, Viale A, Leiner I, Equinda M, Khanin R, Pamer EG: **Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice.** *Journal of Experimental Medicine* 2012, **209**:1445–1456.

93. Fukata M, Chen A, Vamadevan AS, Cohen J, Breglio K, Krishnareddy S, Hsu D, Xu R, Harpaz N, Dannenberg AJ, et al.: **Toll-like receptor-4 promotes the development of colitis-associated colorectal tumors.** *Gastroenterology* 2007, **133**:1869–1881.

94. Fukata M, Shang L, Santaolalla R, Sotolongo J, Pastorini C, España C, Ungaro R, Harpaz N, Cooper HS, Elson G, et al.: **Constitutive activation of epithelial TLR4 augments inflammatory responses to mucosal injury and drives colitis-associated tumorigenesis.** *Inflamm. Bowel Dis.* 2011, **17**:1464–1473.

95. Lowe EL, Crother TR, Rabizadeh S, Hu B, Wang H, Chen S, Shimada K, Wong MH, Michelsen KS, Arditi M: **Toll-like receptor 2 signaling protects mice from tumor development in a mouse model of colitis-induced cancer.** *PLoS ONE* 2010, **5**:e13027.

96. Mariathasan S, Monack DM: **Inflammasome adaptors and sensors: intracellular regulators of infection and inflammation**. *Nat. Rev. Immunol.* 2007, **7**:31–40.

97. McIntire CR, Yeretssian G, Saleh M: **Inflammasomes in infection and inflammation**. *Apoptosis* 2009, **14**:522–535.

98. Gross O, Thomas CJ, Guarda G, Tschopp J: **The inflammasome: an integrated view.** *Immunol. Rev.* 2011, **243**:136–151.

99. Franchi L, Muñoz-Planillo R, Núñez G: **Sensing and reacting to microbes through the inflammasomes**. *Nat. Immunol.* 2012, **13**:325–332.

100. Petnicki-Ocwieja T, Hrncir T, Liu Y-J, Biswas A, Hudcovic T, Tlaskalova-Hogenova H, Kobayashi KS: **Nod2 is required for the regulation of commensal microbiota in the intestine.** *Proceedings of the National Academy of Sciences* 2009, **106**:15813–15818.

101. Natividad JMM, Petit V, Huang X, de Palma G, Jury J, Sanz Y, Philpott D, Garcia Rodenas CL, McCoy KD, Verdu EF: **Commensal and probiotic bacteria influence intestinal barrier function and susceptibility to colitis in Nod1-/-; Nod2-/- mice.** *Inflamm. Bowel Dis.* 2012, **18**:1434–1446.

102. Chamaillard M, Hashimoto M, Horie Y, Masumoto J, Qiu S, Saab L, Ogura Y, Kawasaki A, Fukase K, Kusumoto S, et al.: **An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid.** *Nat. Immunol.* 2003, **4**:702–707.

103. Girardin SE, Boneca IG, Carneiro LAM, Antignac A, Jéhanno M, Viala J, Tedin K, Taha M-K, Labigne A, Zähringer U, et al.: **Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan.** *Science* 2003, **300**:1584–1587.

104. Chen GY, Shaw MH, Redondo G, Núñez G: **The innate immune receptor Nod1 protects the intestine from inflammation-induced tumorigenesis.** *Cancer Res.* 2008, **68**:10060–10067.

105. Elinav E, Strowig T, Kau AL, Henao-Mejia J, Thaiss CA, Booth CJ, Peaper DR, Bertin J, Eisenbarth SC, Gordon JI, et al.: **NLRP6 Inflammasome Regulates Colonic Microbial Ecology and Risk for Colitis**. *Cell* 2011, **145**:745–757.

106. Elinav E, Strowig T, Henao-Mejia J, Flavell RA: **Regulation of the Antimicrobial Response by NLR Proteins**. *Immunity* 2011, **34**:665–679.

107. Anand PK, Malireddi RKS, Lukens JR, Vogel P, Bertin J, Lamkanfi M, Kanneganti T-D: **NLRP6 negatively regulates innate immunity and host defence against bacterial pathogens**. *Nature* 2012, doi:10.1038/nature11250.

108. Chen GY, Liu M, Wang F, Bertin J, Núñez G: **A functional role for Nlrp6 in intestinal inflammation and tumorigenesis.** *J. Immunol.* 2011, **186**:7187–7194.

109. Normand S, Delanoye-Crespin A, Bressenot A, Huot L, Grandjean T, Peyrin-Biroulet L, Lemoine Y, Hot D, Chamaillard M: **Nod-like receptor pyrin domain-containing protein 6 (NLRP6) controls epithelial self-renewal and colorectal carcinogenesis upon injury.** *Proceedings of the National Academy of Sciences* 2011, **108**:9601–9606.

110. Zaki MH, Boyd KL, Vogel P, Kastan MB, Lamkanfi M, Kanneganti T-D: **The NLRP3 inflammasome protects against loss of epithelial integrity and mortality during experimental colitis.** *Immunity* 2010, **32**:379–391.

111. Allen IC, TeKippe EM, Woodford R-MT, Uronis JM, Holl EK, Rogers AB, Herfarth HH, Jobin C, Ting JPY: **The NLRP3 inflammasome functions as a negative regulator of tumorigenesis during colitis-associated cancer**. *J. Exp. Med.* 2010, **207**:1045–1056.

112. Zaki MH, Vogel P, Malireddi RKS, Body-Malapel M, Anand PK, Bertin J, Green DR, Lamkanfi M, Kanneganti T-D: **The NOD-Like Receptor NLRP12 Attenuates Colon Inflammation and Tumorigenesis**. *Cancer Cell* 2011, **20**:649–660.

113. Allen IC, Wilson JE, Schneider M, Lich JD, Roberts RA, Arthur JC, Woodford R-MT, Davis BK, Uronis JM, Herfarth HH, et al.: **NLRP12 suppresses colon inflammation and tumorigenesis through the**

negative regulation of noncanonical NF-κB signaling. *Immunity* 2012, **36**:742–754.

114. Brinkman BM, Hildebrand F, Kubica M, Goosens D, Del Favero J, Declercq W, Raes J, Vandenabeele P: **Caspase deficiency alters the murine gut microbiome**. *ISME J* 2011, **2**:e220.

115. Hu B, Elinav E, Huber S, Booth CJ, Strowig T, Jin C, Eisenbarth SC, Flavell RA: **Inflammation-induced tumorigenesis in the colon is regulated by caspase-1 and NLRC4.** *Proceedings of the National Academy of Sciences* 2010, **107**:21635–21640.

116. Dupaul-Chicoine J, Yeretssian G, Doiron K, Bergstrom KSB, McIntire CR, LeBlanc PM, Meunier C, Turbide C, Gros P, Beauchemin N, et al.: **Control of intestinal homeostasis, colitis, and colitis-associated colorectal cancer by the inflammatory caspases.** *Immunity* 2010, **32**:367–378.

117. Froy O: **Regulation of mammalian defensin expression by Toll-like receptor-dependent and independent signalling pathways**. *Cell. Microbiol.* 2005, **7**:1387–1397.

118. Salzman NH, Hung K, Haribhai D, Chu H, Karlsson-Sjöberg J, Amir E, Teggatz P, Barman M, Hayward M, Eastwood D, et al.: **Enteric defensins are essential regulators of intestinal microbial ecology.** *Nat. Immunol.* 2010, **11**:76–83.

119. Tanabe H, Sato T, Watari J, Maemoto A, Fijiya M, Kono T, Ashida T, Ayabe T, Kohgo Y: **Functional Role of Metaplastic Paneth Cell Defensins in Helicobacter pylori-Infected Stomach**. *Helicobacter* 2008, **13**:370–379.

120. Grivennikov SI, Karin M: **Inflammation and oncogenesis: a vicious connection.** *Curr. Opin. Genet. Dev.* 2010, **20**:65–71.

121. Grivennikov SI, Karin M: **Dangerous liaisons: STAT3 and NF-kappaB collaboration and crosstalk in cancer.** *Cytokine Growth Factor Rev.* 2010, **21**:11–19.

122. Greten FR, Eckmann L, Greten TF, Park JM, Li Z-W, Egan LJ, Kagnoff MF, Karin M: **IKKbeta links inflammation and tumorigenesis in a mouse model of colitis-associated cancer.** *Cell* 2004, **118**:285–296.

123. Bollrath J, Phesse TJ, Burstin von VA, Putoczki T, Bennecke M, Bateman T, Nebelsiek T, Lundgren-May T, Canli O, Schwitalla S, et al.: **gp130-mediated Stat3 activation in enterocytes regulates cell survival and cell-cycle progression during colitis-associated tumorigenesis.** *Cancer Cell* 2009, **15**:91–102.

124. Grivennikov S, Karin E, Terzić J, Mucida D, Yu G-Y, Vallabhapurapu S, Scheller J, Rose-John S, Cheroutre H, Eckmann L: **IL-6 and Stat3 Are Required for Survival of Intestinal Epithelial Cells and Development of Colitis-Associated Cancer**. *Cancer Cell* 2009, **15**:103–113.

125. Wang S, Liu Z, Wang L, Zhang X: **NF-kappaB signaling pathway, inflammation and colorectal cancer.** *Cell. Mol. Immunol.* 2009, **6**:327–334.

126. Popivanova BK, Kitamura K, Wu Y, Kondo T, Kagaya T, Kaneko S, Oshima M, Fujii C, Mukaida N: **Blocking TNF-alpha in mice reduces colorectal carcinogenesis associated with chronic colitis.** *J. Clin. Invest.* 2008, **118**:560–570.

127. Becker C, Fantini MC, Schramm C, Lehr HA, Wirtz S, Nikolaev A, Burg J, Strand S, Kiesslich R, Huber S, et al.: **TGF-beta suppresses tumor progression in colon cancer by inhibition of IL-6 trans-signaling.** *Immunity* 2004, **21**:491–501.

128. Becker C, Fantini MC, Wirtz S, Nikolaev A, Lehr HA, Galle PR, Rose-John S, Neurath MF: **IL-6 signaling promotes tumor growth in colorectal cancer.** *Cell Cycle* 2005, **4**:217–220.

129. Wang L, Yi T, Kortylewski M, Pardoll DM, Zeng D, Yu H: **IL-17 can promote tumor growth through an IL-6–Stat3 signaling pathway**. *J. Exp. Med.* 2009, **206**:1457–1464.

130. Dominitzki S, Fantini MC, Neufert C, Nikolaev A, Galle PR, Scheller J, Monteleone G, Rose-John S, Neurath MF, Becker C: **Cutting edge: trans-signaling via the soluble IL-6R abrogates the induction of FoxP3 in naive CD4+CD25 T cells.** *J. Immunol.* 2007, **179**:2041–2045.

131. Bettelli E, Carrier Y, Gao W, Korn T, Strom TB, Oukka M, Weiner HL, Kuchroo VK: **Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells.** *Nature* 2006, **441**:235–238.

132. Yu H, Pardoll D, Jove R: **STATs in cancer inflammation and immunity: a leading role for STAT3**. *Nat. Rev. Cancer* 2009, **9**:798–809.

133. Langowski JL, Zhang X, Wu L, Mattson JD, Chen T, Smith K, Basham B, McClanahan T, Kastelein RA, Oft M: **IL-23 promotes tumour incidence and growth.** *Nature* 2006, **442**:461–465.

134. Langowski JL, Kastelein RA, Oft M: **Swords into plowshares: IL-23 repurposes tumor immune surveillance.** *Trends Immunol.* 2007, **28**:207–212.

135. Wu S, Rhee K-J, Albesiano E, Rabizadeh S, Wu X, Yen H-R, Huso DL, Brancati FL, Wick E, McAllister F, et al.: **A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses.** *Nat. Med.* 2009, **15**:1016–1022.

136. Grivennikov SI, Wang K, Mucida D, Stewart CA, Schnabl B, Jauch D, Taniguchi K, Yu G-Y, Österreicher CH, Hung KE, et al.: **Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth**. *Nature* 2012, doi:10.1038/nature11465.

137. Kipanyula MJ, Seke Etet PF, Vecchio L, Farahna M, Nukenine EN, Nwabo Kamdje AH: **Signaling pathways bridging microbial-triggered inflammation and cancer**. *Cellular Signalling* 2013, **25**:403–416.

138. Smith WL, DeWitt DL, Garavito RM: **Cyclooxygenases: structural, cellular, and molecular biology.** *Annu Rev Biochem* 2000, **69**:145–182.

139. Eberhart CE, Coffey RJ, Radhika A, Giardiello FM, Ferrenbach S, DuBois RN: **Up-regulation of cyclooxygenase 2 gene expression in human colorectal adenomas and adenocarcinomas.** *Gastroenterology* 1994, **107**:1183–1188.

140. Sano H, Kawahito Y, Wilder RL, Hashiramoto A, Mukai S, Asai K, Kimura S, Kato H, Kondo M, Hla T: **Expression of cyclooxygenase-1 and -2 in human colorectal cancer.** *Cancer Res.* 1995, **55**:3785–3789.

141. Kargman SL, O'Neill GP, Vickers PJ, Evans JF, Mancini JA, Jothy S: **Expression of prostaglandin G/H synthase-1 and -2 protein in human colon cancer.** *Cancer Res.* 1995, **55**:2556–2559.

142. Williams CS, Luongo C, Radhika A, Zhang T, Lamps LW, Nanney LB, Beauchamp RD, DuBois RN: **Elevated cyclooxygenase-2 levels in Min mouse adenomas.** *Gastroenterology* 1996, **111**:1134–1140.

143. Giovannucci E, Rimm EB, Stampfer MJ, Colditz GA, Ascherio A, Willett WC: **Aspirin use and the risk for colorectal cancer and adenoma in male health professionals.** *Ann Intern Med* 1994, **121**:241–246.

144. Giovannucci E, Egan KM, Hunter DJ, Stampfer MJ, Colditz GA, Willett WC, Speizer FE: **Aspirin and the risk of colorectal cancer in women.** *N. Engl. J. Med.* 1995, **333**:609–614.

145. Thun MJ, Namboodiri MM, Heath CWJ: **Aspirin use and reduced risk of fatal colon cancer.** *N. Engl. J.*

Med. 1991, **325**:1593–1596.

146. Gupta RA, DuBois RN: **Colorectal cancer prevention and treatment by inhibition of cyclooxygenase-2.** *Nat. Rev. Cancer* 2001, **1**:11–21.

147. Nugent KP, Farmer KC, Spigelman AD, Williams CB, Phillips RK: **Randomized controlled trial of the effect of sulindac on duodenal and rectal polyposis and cell proliferation in patients with familial adenomatous polyposis.** *Br J Surg* 1993, **80**:1618–1619.

148. Giardiello FM, Hamilton SR, Krush AJ, Piantadosi S, Hylind LM, Celano P, Booker SV, Robinson CR, Offerhaus GJ: **Treatment of colonic and rectal adenomas with sulindac in familial adenomatous polyposis.** *N. Engl. J. Med.* 1993, **328**:1313–1316.

149. Morgan X, Tickle T, Sokol H, Gevers D, Devaney K, Ward D, Reyes J, Shah S, Leleiko N, Snapper S, et al.: **Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment**. *Genome Biol.* 2012, **13**:R79.

150. Nones K, Knoch B, Dommels YEM, Paturi G, Butts C, McNabb WC, Roy NC: **Multidrug resistance gene deficient (mdr1a-/-) mice have an altered caecal microbiota that precedes the onset of intestinal inflammation.** *J. Appl. Microbiol.* 2009, **107**:557–566.

151. Garrett WS, Lord GM, Punit S, Lugo-Villarino G, Mazmanian SK, Ito S, Glickman JN, Glimcher LH: **Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system.** *Cell* 2007, **131**:33–45.

152. Garrett WS, Gallini CA, Yatsunenko T, Michaud M, DuBois A, Delaney ML, Punit S, Karlsson M, Bry L, Glickman JN: **Enterobacteriaceae Act in Concert with the Gut Microbiota to Induce Spontaneous and Maternally Transmitted Colitis**. *Cell Host Microbe* 2010, **8**:292–300.

153. Bloom SM, Bijanki VN, Nava GM, Sun L, Malvin NP, Donermeyer DL, Dunne WM, Allen PM, Stappenbeck TS: **Commensal Bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease.** *Cell Host Microbe* 2011, **9**:390–403.

154. Garrett WS, Gordon JI, Glimcher LH: **Homeostasis and inflammation in the intestine.** *Cell* 2010, **140**:859–870.

155. Garrett WS, Punit S, Gallini CA, Michaud M, Zhang D, Sigrist KS, Lord GM, Glickman JN, Glimcher LH: **Colitis-associated colorectal cancer driven by T-bet deficiency in dendritic cells.** *Cancer Cell* 2009, **16**:208–219.

156. Marshall BJ, Warren JR: **Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration.** *Lancet* 1984, **1**:1311–1315.

157. Parkin DM, Dharmani P, Bray F, Strauss J, Ferlay J, Ambrose C, Pisani P, Allen-Vercoe E, Chadee K: **Global cancer statistics, 2002.** *CA Cancer J Clin* 2005, **55**:74–108.

158. Eslick G-D: **Helicobacter pylori infection causes gastric cancer? A review of the epidemiological, meta-analytic, and experimental evidence. [Internet]**. *World J. Gastroenterol.* 2006, **12**:2991–2999.

159. Huang JQ, Sridhar S, Chen Y, Hunt RH: **Meta-analysis of the relationship between Helicobacter pylori seropositivity and gastric cancer.** *Gastroenterology* 1998, **114**:1169–1179.

160. Eslick GD, Lim LL, Byles JE, Xia HH, Talley NJ: **Association of Helicobacter pylori infection with gastric carcinoma: a meta-analysis.** *Am J Gastroenterol* 1999, **94**:2373–2379.

161. Danesh J: **Helicobacter pylori infection and gastric cancer: systematic review of the epidemiological studies.** *Alimentary Pharmacology & Therapeutics* 1999, **13**:851–856.

162. Xue FB, Xu YY, Wan Y, Pan BR, Ren J, Fan DM: **Association of H. pylori infection with gastric carcinoma: a Meta analysis.** *World J. Gastroenterol.* 2001, **7**:801–804.

163. Axon A: **Review article: gastric cancer and Helicobacter pylori.** *Alimentary Pharmacology & Therapeutics* 2002, **16 Suppl 4**:83–88.

164. Huang JQ, Zheng GF, Sumanac K, Irvine EJ, Hunt RH: **Meta-analysis of the relationship between cagA seropositivity and gastric cancer.** *Gastroenterology* 2003, **125**:1636–1644.

165. Wong BC-Y, Lam SK, Wong WM, Chen JS, Zheng TT, Feng RE, Lai KC, Hu WHC, Yuen ST, Leung SY, et al.: **Helicobacter pylori eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial.** *JAMA* 2004, **291**:187–194.

166. Gisbert JP, Calvet X: **Review article: common misconceptions in the management of Helicobacter pylori-associated gastric MALT-lymphoma.** *Alimentary Pharmacology & Therapeutics* 2011, **34**:1047–1062.

167. Luman W: ***Helicobacter pylori*: causation and treatment**. *J R Coll Physicians Edinb* 2005, **35**:45–49.

168. Galmiche A, Rassow J, Doye A, Cagnol S, Chambard JC, Contamin S, de Thillot V, Just I, Ricci V, Solcia E, et al.: **The N-terminal 34 kDa fragment of Helicobacter pylori vacuolating cytotoxin targets mitochondria and induces cytochrome c release.** *EMBO J* 2000, **19**:6361–6370.

169. Higashi H, Tsutsumi R, Muto S, Sugiyama T, Azuma T, Asaka M, Hatakeyama M: **SHP-2 tyrosine phosphatase as an intracellular target of Helicobacter pylori CagA protein.** *Science* 2002, **295**:683–686.

170. Higashi H, Nakaya A, Tsutsumi R, Yokoyama K, Fujii Y, Ishikawa S, Higuchi M, Takahashi A, Kurashima Y, Teishikata Y, et al.: **Helicobacter pylori CagA induces Ras-independent morphogenetic response through SHP-2 recruitment and activation.** *J. Biol. Chem.* 2004, **279**:17205–17216.

171. Klein RS, Recco RA, Catalano MT, Edberg SC, Casey JI, Steigbigel NH: **Association of Streptococcus bovis with carcinoma of the colon.** *N. Engl. J. Med.* 1977, **297**:800–802.

172. Boleij A, van Gelder MMHJ, Swinkels DW, Tjalsma H: **Clinical Importance of Streptococcus gallolyticus Infection Among Colorectal Cancer Patients: Systematic Review and Meta-analysis**. *Clinical Infectious Diseases* 2011, **53**:870–878.

173. Boleij A, Schaeps RMJ, Tjalsma H: **Association between Streptococcus bovis and colon cancer.** *J. Clin. Microbiol.* 2009, **47**:516.

174. Gupta A, Madani R, Mukhtar H: **Streptococcus bovisendocarditis, a silent sign for colonic tumour**. *J. Surg. Oncol.* 2010, **12**:164–171.

175. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE: **A bacterial driver–passenger model for colorectal cancer: beyond the usual suspects**. *Nature Reviews Microbiology* 2012, **10**:575–582.

176. Nougayrède J-P, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, Buchrieser C, Hacker J, Dobrindt U, Oswald E: **Escherichia coli induces DNA double-strand breaks in eukaryotic cells.** *Science* 2006, **313**:848–851.

177. Cuevas-Ramos G, Petit CR, Marcq I, Boury M, Oswald E, Nougayrède J-P: **Escherichia coli induces**

DNA damage in vivo and triggers genomic instability in mammalian cells. *Proceedings of the National Academy of Sciences* 2010, **107**:11537–11542.

178. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan T-J, Campbell BJ, Abujamel T, Dogan B, Rogers AB, et al.: **Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota**. *Science* 2012, **338**:120–123.

179. Swidsinski A, Khilkin M, Kerjaschki D, Schreiber S, Ortner M, Weber J, Lochs H: **Association between intraepithelial Escherichia coli and colorectal cancer**. *Gastroenterology* 1998, **115**:281–286.

180. Martin HM, Campbell BJ, Hart CA, Mpofu C, Nayar M, Singh R, Englyst H, Williams HF, Rhodes JM: **Enhanced Escherichia coli adherence and invasion in Crohn's disease and colon cancer**. *Gastroenterology* 2004, **127**:80–93.

181. Sears CL: **Enterotoxigenic Bacteroides fragilis: a Rogue among Symbiotes**. *Clin. Microbiol. Rev.* 2009, **22**:349–369.

182. Toprak NU, Yagci A, Gulluoglu BM, Akin ML, Demirkalem P, Celenk T, Soyletir G: **A possible role of Bacteroides fragilis enterotoxin in the aetiology of colorectal cancer.** *Clin. Microbiol. Infect.* 2006, **12**:782–786.

183. Wu S, Morin PJ, Maouyo D, Sears CL: **Bacteroides fragilis enterotoxin induces c-Myc expression and cellular proliferation.** *Gastroenterology* 2003, **124**:392–400.

184. Rabizadeh S, Rhee K-J, Wu S, Huso D, Gan CM, Golub JE, Wu X, Zhang M, Sears CL: **Enterotoxigenic bacteroides fragilis: a potential instigator of colitis.** *Inflamm. Bowel Dis.* 2007, **13**:1475–1483.

185. Rhee K-J, Wu S, Wu X, Huso DL, Karim B, Franco AA, Rabizadeh S, Golub JE, Mathews LE, Shin J, et al.: **Induction of persistent colitis by a human commensal, enterotoxigenic Bacteroides fragilis, in wild-type C57BL/6 mice.** *Infection and Immunity* 2009, **77**:1708–1718.

186. Goodwin AC, Destefano Shields CE, Wu S, Huso DL, Wu X, Murray-Stewart TR, Hacker-Prietz A, Rabizadeh S, Woster PM, Sears CL, et al.: **Polyamine catabolism contributes to enterotoxigenic Bacteroides fragilis-induced colon tumorigenesis.** *Proceedings of the National Academy of Sciences* 2011, **108**:15354–15359.

187. Sears CL, Pardoll DM: **Perspective: Alpha-Bugs, Their Microbial Partners, and the Link to Colon Cancer**. *Journal of Infectious Diseases* 2011, **203**:306–311.

188. Ward TG, Trexler PC: **Gnotobiotics: a new discipline in biological and medical research.** *Perspect. Biol. Med.* 1958, **1**:447–456.

189. Yi P, Li L: **The germfree murine animal: an important animal model for research on the relationship between gut microbiota and the host.** *Vet. Microbiol.* 2012, **157**:1–7.

190. Cash HL, Whitham CV, Behrendt CL, Hooper LV: **Symbiotic bacteria direct expression of an intestinal bactericidal lectin.** *Science* 2006, **313**:1126–1130.

191. Yamamoto M, Yamaguchi R, Munakata K, Takashima K, Nishiyama M, Hioki K, Ohnishi Y, Nagasaki M, Imoto S, Miyano S, et al.: **A microarray analysis of gnotobiotic mice indicating that microbial exposure during the neonatal period plays an essential role in immune system development.** *BMC Genomics* 2012, **13**:335.

192. Dalmasso G, Nguyen HTT, Yan Y, Laroui H, Charania MA, Ayyadurai S, Sitaraman SV, Merlin D: **Microbiota modulate host gene expression via microRNAs.** *PLoS ONE* 2011, **6**:e19293.

193. Reddy BS, Mastromarino A, Wynder EL: **Further leads on metabolic epidemiology of large bowel cancer.** *Cancer Res.* 1975, **35**:3403–3406.

194. Kado S, Uchida K, Funabashi H, Iwata S, Nagata Y, Ando M, Onoue M, Matsuoka Y, Ohwaki M, Morotomi M: **Intestinal microflora are necessary for development of spontaneous adenocarcinoma of the large intestine in T-cell receptor beta chain and p53 double-knockout mice. [Internet]**. *Cancer Res.* 2001, **61**:2395–2398.

195. Engle SJ, Ormsby I, Pawlowski S, Boivin GP, Croft J, Balish E, Doetschman T: **Elimination of colon cancer in germ-free transforming growth factor beta 1-deficient mice.** *Cancer Res.* 2002, **62**:6362–6366.

196. Erdman SE, Poutahidis T, Tomczak M, Rogers AB, Cormier K, Plank B, Horwitz BH, Fox JG: **CD4+ CD25+ regulatory T lymphocytes inhibit microbially induced colon cancer in Rag2-deficient mice.** *Am. J. Pathol.* 2003, **162**:691–702.

197. Li Y, Kundu P, Seow SW, de Matos CT, Aronsson L, Chin KC, Karre K, Pettersson S, Greicius G: **Gut microbiota accelerate tumor growth via c-jun and STAT3 phosphorylation in APCMin/+ mice**. *Carcinogenesis* 2012, **33**:1231–1238.

198. Kühn R, Löhler J, Rennick D, Rajewsky K, Müller W: **Interleukin-10-deficient mice develop chronic enterocolitis**. *Cell* 1993, **75**:263–274.

199. Moore KW, O'Garra A, Malefyt RW, Vieira P, Mosmann TR: **Interleukin-10**. *Annu. Rev. Immunol.* 1993, **11**:165–190.

200. Balish E, Warner T: **Enterococcus faecalis Induces Inflammatory Bowel Disease in Interleukin-10 Knockout Mice**. *Am. J. Pathol.* 2002, **160**:2253–2257.

201. Uronis JM, Mühlbauer M, Herfarth HH, Rubinas TC, Jones GS, Jobin C: **Modulation of the intestinal microbiota alters colitis-associated colorectal cancer susceptibility.** *PLoS ONE* 2009, **4**:e6026.

202. Nagamine CM, Rogers AB, Fox JG, Schauer DB: **Helicobacter hepaticus promotes azoxymethane-initiated colon tumorigenesis in BALB/c-IL10-deficient mice**. *Int J Cancer* 2007, **122**:832–838.

203. Boulard O, Kirchberger S, Royston DJ, Maloy KJ, Powrie FM: **Identification of a genetic locus controlling bacteria-driven colitis and associated cancer through effects on innate inflammation**. *J. Exp. Med.* 2012, **209**:1309–1324.

204. Laqueur GL, McDaniel EG, Matsumoto H: **Tumor induction in germfree rats with methylazoxymethanol (MAM) and synthetic MAM acetate.** *J. Natl. Cancer Inst.* 1967, **39**:355–371.

205. Rowland IR: **The role of the gastrointestinal microbiota in colorectal cancer.** *Curr. Pharm. Des.* 2009, **15**:1524–1527.

206. Fiala ES: **Investigations into the metabolism and mode of action of the colon carcinogens 1,2-dimethylhydrazine and azoxymethane.** *Cancer* 1977, **40**:2436–2445.

207. Neufert C, Becker C, Neurath MF: **An inducible mouse model of colon carcinogenesis for the analysis of sporadic and inflammation-driven tumor progression.** *Nat Protoc* 2007, **2**:1998–2004.

208. Takada H, Hirooka T, Hiramatsu Y, Yamamoto M: **Effect of beta-glucuronidase inhibitor on azoxymethane-induced colonic carcinogenesis in rats.** *Cancer Res.* 1982, **42**:331–334.

209. Wallace BD, Wang H, Lane KT, Scott JE, Orans J, Koo JS, Venkatesh M, Jobin C, Yeh L-A, Mani S, et

al.: **Alleviating cancer drug toxicity by inhibiting a bacterial enzyme.** *Science* 2010, **330**:831–835.

210. Kim D-H, Jin Y-H: **Intestinal bacterial β-glucuronidase activity of patients with colon cancer**. *Archives of Pharmaceutical Research* 2001, **24**:564–567.

211. Reddy BS, Mangat S, Weisburger JH, Wynder EL: **Effect of High-Risk Diets for Colon Carcinogenesis on Intestinal Mucosal and Bacterial β-Glucuronidase Activity in F344 Rats**. *Cancer Res.* 1977, **37**:3533–3536.

212. Mazière S, Meflah K, Tavan E, Champ M, Narbonne JF, Cassand P: **Effect of resistant starch and/or fat-soluble vitamins A and E on the initiation stage of aberrant crypts in rat colon**. *Nutr Cancer* 1998, **31**:168–177.

213. Gråsten SM, Juntunen KS, Poutanen KS, Gylling HK, Miettinen TA, Mykkänen HM: **Rye bread improves bowel function and decreases the concentrations of some compounds that are putative colon cancer risk markers in middle-aged women and men.** *J. Nutr.* 2000, **130**:2215–2221.

214. Powolny A, Xu J, Loo G: **Deoxycholate induces DNA damage and apoptosis in human colon epithelial cells expressing either mutant or wild-type p53.** *Int. J. Biochem. Cell Biol.* 2001, **33**:193–203.

215. Owen RW: **Faecal steroids and colorectal carcinogenesis.** *Scand. J. Gastroenterol. Suppl.* 1997, **222**:76–82.

216. Gill CIR, Rowland IR: **Diet and cancer: assessing the risk.** *Br. J. Nutr.* 2002, **88 Suppl 1**:S73–87.

217. Huycke MM, Abrams V, Moore DR: **Enterococcus faecalis produces extracellular superoxide and hydrogen peroxide that damages colonic epithelial cell DNA.** *Carcinogenesis* 2002, **23**:529–536.

218. Wang X, Huycke MM: **Extracellular superoxide production by Enterococcus faecalis promotes chromosomal instability in mammalian cells.** *Gastroenterology* 2007, **132**:551–561.

219. Wang X, Allen TD, May RJ, Lightfoot S, Houchen CW, Huycke MM: **Enterococcus faecalis induces aneuploidy and tetraploidy in colonic epithelial cells through a bystander effect.** *Cancer Res.* 2008, **68**:9909–9917.

220. Meira LB, Bugni JM, Green SL, Lee C-W, Pang B, Borenshtein D, Rickman BH, Rogers AB, Moroski-Erkul CA, McFaline JL, et al.: **DNA damage induced by chronic inflammation contributes to colon carcinogenesis in mice.** *J. Clin. Invest.* 2008, **118**:2516–2525.

221. Hussain SP, Hofseth LJ, Harris CC: **Radical causes of cancer**. *Nat. Rev. Cancer* 2003, **3**:276–285.

222. Mangerich A, Knutson CG, Parry NM, Muthupalani S, Ye W, Prestwich E, Cui L, McFaline JL, Mobley M, Ge Z, et al.: **Infection-induced colitis in mice causes dynamic and tissue-specific changes in stress response and DNA damage leading to colon cancer.** *Proceedings of the National Academy of Sciences* 2012, doi:10.1073/pnas.1207829109.

223. Shaked H, Hofseth LJ, Chumanevich A, Chumanevich AA, Wang J, Wang Y, Taniguchi K, Guma M, Shenouda S, Clevers H, et al.: **Chronic epithelial NF-κB activation accelerates APC loss and intestinal tumor initiation through iNOS up-regulation.** *Proceedings of the National Academy of Sciences* 2012, **109**:14007–14012.

224. Goodman JE, Hofseth LJ, Hussain SP, Harris CC: **Nitric oxide and p53 in cancer-prone chronic inflammation and oxyradical overload disease.** *Environ. Mol. Mutagen.* 2004, **44**:3–9.

225. Erdman SE, Rao VP, Poutahidis T, Rogers AB, Taylor CL, Jackson EA, Ge Z, Lee CW, Schauer DB, Wogan GN, et al.: **Nitric oxide and TNF-alpha trigger colonic inflammation and carcinogenesis in Helicobacter hepaticus-infected, Rag2-deficient mice.** *Proceedings of the National Academy of Sciences* 2009, **106**:1027–1032.

226. Chu FF: **Bacteria-Induced Intestinal Cancer in Mice with Disrupted Gpx1 and Gpx2 Genes**. *Cancer Res.* 2004, **64**:962–968.

227. Wink DA, Kasprzak KS, Maragos CM, Elespuru RK, Misra M, Dunams TM, Cebula TA, Koch WH, Andrews AW, Allen JS: **DNA deaminating ability and genotoxicity of nitric oxide and its progenitors.** *Science* 1991, **254**:1001–1003.

228. Nguyen T, Brunson D, Crespi CL, Penman BW, Wishnok JS, Tannenbaum SR: **DNA damage and mutation in human cells exposed to nitric oxide in vitro.** *Proc. Natl. Acad. Sci. U.S.A.* 1992, **89**:3030–3034.

229. Massey RC, Key PE, Mallett AK, Rowland IR: **An investigation of the endogenous formation of apparent total N-nitroso compounds in conventional microflora and germ-free rats.** *Food Chem Toxicol* 1988, **26**:595–600.

230. Zhang Z-W, Abdullahi M, Farthing MJG: **Effect of physiological concentrations of vitamin C on gastric cancer cells and Helicobacter pylori.** *Gut* 2002, **50**:165–169.

231. Humblot C, Murkovic M, Rigottier-Gois L, Bensaada M, Bouclet A, Andrieux C, Anba J, Rabot S: **Beta-glucuronidase in human intestinal microbiota is necessary for the colonic genotoxicity of the food-borne carcinogen 2-amino-3-methylimidazo[4,5-f]quinoline in rats.** *Carcinogenesis* 2007, **28**:2419–2425.

232. Vanhaecke L, Vercruysse F, Boon N, Verstraete W, Cleenwerck I, De Wachter M, De Vos P, van de Wiele T: **Isolation and characterization of human intestinal bacteria capable of transforming the dietary carcinogen 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.** *Appl. Environ. Microbiol.* 2008, **74**:1469–1477.

233. Kassie F, Rabot S, Kundi M, Chabicovsky M, Qin HM, Knasmüller S: **Intestinal microflora plays a crucial role in the genotoxicity of the cooked food mutagen 2-amino-3-methylimidazo [4,5-f]quinoline.** *Carcinogenesis* 2001, **22**:1721–1725.

234. Knasmüller S, Steinkellner H, Hirschl AM, Rabot S, Nobis EC, Kassie F: **Impact of bacteria in dairy products and of the intestinal microflora on the genotoxic and carcinogenic effects of heterocyclic aromatic amines.** *Mutat. Res.* 2001, **480-481**:129–138.

235. Bagnoli F, Buti L, Tompkins L, Covacci A, Amieva MR: **Helicobacter pylori CagA induces a transition from polarized to invasive phenotypes in MDCK cells.** *Proc. Natl. Acad. Sci. U.S.A.* 2005, **102**:16339–16344.

236. Segal ED, Cha J, Lo J, Falkow S, Tompkins LS: **Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by Helicobacter pylori.** *Proc. Natl. Acad. Sci. U.S.A.* 1999, **96**:14559–14564.

237. Peek RM, Vaezi MF, Falk GW, Goldblum JR, Perez-Perez GI, Richter JE, Blaser MJ: **Role ofHelicobacter pylori cagA+ strains and specific host immune responses on the development of premalignant and malignant lesions in the gastric cardia**. *Int J Cancer* 1999, **82**:520–524.

238. Urbanska AM, Bhathena J, Martoni C, Prakash S: **Estimation of the potential antitumor activity of microencapsulated Lactobacillus acidophilus yogurt formulation in the attenuation of tumorigenesis**

in Apc(Min/+) mice. *Dig. Dis. Sci.* 2009, **54**:264–273.

239. Arimochi H, Kinouchi T, Kataoka K, Kuwahara T, Ohnishi Y: **Effect of intestinal bacteria on formation of azoxymethane-induced aberrant crypt foci in the rat colon.** *Biochem. Biophys. Res. Commun.* 1997, **238**:753–757.

240. Chang J-H, Shim YY, Cha S-K, Reaney MJT, Chee KM: **Effect of Lactobacillus acidophilus KFRI342 on the development of chemically induced precancerous growths in the rat colon.** *J. Med. Microbiol.* 2012, **61**:361–368.

241. Foo N-P, Ou Yang H, Chiu H-H, Chan H-Y, Liao C-C, Yu C-K, Wang Y-J: **Probiotics prevent the development of 1,2-dimethylhydrazine (DMH)-induced colonic tumorigenesis through suppressed colonic mucosa cellular proliferation and increased stimulation of macrophages.** *J. Agric. Food Chem.* 2011, **59**:13337–13345.

242. Roller M, Pietro Femia A, Caderni G, Rechkemmer G, Watzl B: **Intestinal immunity of rats with colon cancer is modulated by oligofructose-enriched inulin combined with Lactobacillus rhamnosus and Bifidobacterium lactis.** *Br. J. Nutr.* 2004, **92**:931–938.

243. Rowland IR, Rumney CJ, Coutts JT, Lievense LC: **Effect of Bifidobacterium longum and inulin on gut bacterial metabolism and carcinogen-induced aberrant crypt foci in rats.** *Carcinogenesis* 1998, **19**:281–285.

244. Goldin BR, Gorbach SL: **Alterations of the intestinal microflora by diet, oral antibiotics, and Lactobacillus: decreased production of free amines from aromatic nitro compounds, azo dyes, and glucuronides.** *J. Natl. Cancer Inst.* 1984, **73**:689–695.

245. Goldin BR, Swenson L, Dwyer J, Sexton M, Gorbach SL: **Effect of diet and Lactobacillus acidophilus supplements on human fecal bacterial enzymes.** *J. Natl. Cancer Inst.* 1980, **64**:255–261.

246. Pool-Zobel BL, Bertram B, Knoll M, Lambertz R, Neudecker C, Schillinger U, Schmezer P, Holzapfel WH: **Antigenotoxic properties of lactic acid bacteria in vivo in the gastrointestinal tract of rats.** *Nutr Cancer* 1993, **20**:271–281.

247. Nowak A, Libudzisz Z: **Ability of probiotic Lactobacillus casei DN 114001 to bind or/and metabolise heterocyclic aromatic amines in vitro.** *Eur J Nutr* 2009, **48**:419–427.

248. Zsivkovits M, Fekadu K, Sontag G, Nabinger U, Huber WW, Kundi M, Chakraborty A, Foissy H, Knasmüller S: **Prevention of heterocyclic amine-induced DNA damage in colon and liver of rats by different lactobacillus strains.** *Carcinogenesis* 2003, **24**:1913–1918.

249. Reddy BS, Rivenson A: **Inhibitory effect of Bifidobacterium longum on colon, mammary, and liver carcinogenesis induced by 2-amino-3-methylimidazo[4,5-f]quinoline, a food mutagen.** *Cancer Res.* 1993, **53**:3914–3918.

250. Uronis JM, Arthur JC, Keku T, Fodor A, Carroll IM, Cruz ML, Appleyard CB, Jobin C: **Gut microbial diversity is reduced by the probiotic VSL#3 and correlates with decreased TNBS-induced colitis**. *Inflamm. Bowel Dis.* 2011, **17**:289–297.

251. Appleyard CB, Cruz ML, Isidro AA, Arthur JC, Jobin C, De Simone C: **Pretreatment with the probiotic VSL#3 delays transition from inflammation to dysplasia in a rat model of colitis-associated cancer**. *Am. J. Physiol. Gastrointest. Liver Physiol.* 2011, **301**:G1004–G1013.

252. Rafter J, Bennett M, Caderni G, Clune Y, Hughes R, Karlsson PC, Klinder A, O'Riordan M, O'Sullivan GC, Pool-Zobel B, et al.: **Dietary synbiotics reduce cancer risk factors in polypectomized and colon**

cancer patients. *Am. J. Clin. Nutr.* 2007, **85**:488–496.

253. Kim Y, Lee D, Kim D, Cho J, Yang J, Chung M, Kim K, Ha N: **Inhibition of proliferation in colon cancer cell lines and harmful enzyme activity of colon bacteria by Bifidobacterium adolescentis SPM0212.** *Arch. Pharm. Res.* 2008, **31**:468–473.

254. Lee N-K, Park J-S, Park E, Paik H-D: **Adherence and anticarcinogenic effects of Bacillus polyfermenticus SCD in the large intestine.** *Lett. Appl. Microbiol.* 2007, **44**:274–278.

255. Baricault L, Denariaz G, Houri JJ, Bouley C, Sapin C, Trugnan G: **Use of HT-29, a cultured human colon cancer cell line, to study the effect of fermented milks on colon cancer cell growth and differentiation.** *Carcinogenesis* 1995, **16**:245–252.

256. Biffi A, Coradini D, Larsen R, Riva L, Di Fronzo G: **Antiproliferative effect of fermented milk on the growth of a human breast cancer cell line.** *Nutr Cancer* 1997, **28**:93–99.

257. Corthésy B, Gaskins HR, Mercenier A: **Cross-talk between probiotic bacteria and the host immune system.** *J. Nutr.* 2007, **137**:781S–90S.

258. Geier MS, Butler RN, Howarth GS: **Probiotics, prebiotics and synbiotics: a role in chemoprevention for colorectal cancer?** *Cancer Biol. Ther.* 2006, **5**:1265–1269.

259. Scheppach W: **Effects of short chain fatty acids on gut morphology and function.** *Gut* 1994, **35**:S35–8.

260. Maslowski KM, Vieira AT, Ng A, Kranich J, Sierro F, Yu D, Schilter HC, Rolph MS, Mackay F, Artis D, et al.: **Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43.** *Nature* 2009, **461**:1282–1286.

261. Heerdt BG, Houston MA, Augenlicht LH: **Short-chain fatty acid-initiated cell cycle arrest and apoptosis of colonic epithelial cells is linked to mitochondrial function.** *Cell Growth Differ.* 1997, **8**:523–532.

262. Bonnotte B, Favre N, Reveneau S, Micheau O, Droin N, Garrido C, Fontana A, Chauffert B, Solary E, Martin F: **Cancer cell sensitization to fas-mediated apoptosis by sodium butyrate.** *Cell Death Differ.* 1998, **5**:480–487.

263. Ruemmele FM, Schwartz S, Seidman EG, Dionne S, Levy E, Lentze MJ: **Butyrate induced Caco-2 cell apoptosis is mediated via the mitochondrial pathway.** *Gut* 2003, **52**:94–100.

264. Riggs MG, Whittaker RG, Neumann JR, Ingram VM: **n-Butyrate causes histone modification in HeLa and Friend erythroleukaemia cells.** *Nature* 1977, **268**:462–464.

265. Yoshida M, Kijima M, Akita M, Beppu T: **Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A.** *J. Biol. Chem.* 1990, **265**:17174–17179.

266. Richon VM, Emiliani S, Verdin E, Webb Y, Breslow R, Rifkind RA, Marks PA: **A class of hybrid polar inducers of transformed cell differentiation inhibits histone deacetylases.** *Proc. Natl. Acad. Sci. U.S.A.* 1998, **95**:3003–3007.

267. Bolden JE, Peart MJ, Johnstone RW: **Anticancer activities of histone deacetylase inhibitors.** *Nat Rev Drug Discov* 2006, **5**:769–784.

268. Rowland IR, Bearne CA, Fischer R, Pool-Zobel BL: **The effect of lactulose on DNA damage induced by DMH in the colon of human flora-associated rats.** *Nutr Cancer* 1996, **26**:37–47.

269. McIntosh GH, Royle PJ, Playne MJ: **A probiotic strain of L. acidophilus reduces DMH-induced large intestinal tumors in male Sprague-Dawley rats.** *Nutr Cancer* 1999, **35**:153–159.

270. Tlaskalova-Hogenova H, Stěpánková R, Kozáková H, Hudcovic T, Vannucci L, Tučková L, Rossmann P, Hrncir T, Kverka M, Zákostelská Z, et al.: **The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases.** *Cell. Mol. Immunol.* 2011, **8**:110–120.

271. Van der Sluis M, De Koning BAE, De Bruijn ACJM, Velcich A, Meijerink JPP, Van Goudoever JB, Büller HA, Dekker J, Van Seuningen I, Renes IB, et al.: **Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection.** *Gastroenterology* 2006, **131**:117–129.

272. Petersson J, Schreiber O, Hansson GC, Gendler SJ, Velcich A, Lundberg JO, Roos S, Holm L, Phillipson M: **Importance and regulation of the colonic mucus barrier in a mouse model of colitis.** *Am. J. Physiol. Gastrointest. Liver Physiol.* 2011, **300**:G327–33.

273. Velcich A, Yang W, Heyer J, Fragale A, Nicholas C, Viani S, Kucherlapati R, Lipkin M, Yang K, Augenlicht L: **Colorectal cancer in mice genetically deficient in the mucin Muc2.** *Science* 2002, **295**:1726–1729.

274. Yang K, Popova NV, Yang WC, Lozonschi I, Tadesse S, Kent S, Bancroft L, Matise I, Cormier RT, Scherer SJ, et al.: **Interaction of Muc2 and Apc on Wnt signaling and in intestinal tumorigenesis: potential role of chronic inflammation.** *Cancer Res.* 2008, **68**:7313–7322.

This page intentionally left blank

CHAPTER 1

PathSeq: software to identify or discover microbes by deep sequencing of human tissue

This Chapter is a reproduction of a published manuscript:
Kostic, A.D., Ojesina, A.I., Pedamallu, C.S., Jung, J., Verhaak, R.G.W., Getz, G., and Meyerson, M. (2011). PathSeq: software to identify or discover microbes by deep sequencing of human tissue. Nat. Biotechnol. 29, 393–396.

Author Contributions:

A.D.K., A.I.O., J.J., R.G.W.V., G.G. and M.M. conceived the study. A.D.K. implemented the method and C.S.P. made it available in a cloud-computing environment. A.D.K. performed the RNA-Seq. A.D.K., A.I.O., C.S.P., J.J., G.G. and M.M. designed and A.D.K. and C.S.P. implemented simulation and validation experiments. A.D.K., C.S.P. and M.M. wrote the manuscript.

Abstract

Many human diseases are believed to be caused by undiscovered pathogens [1,2]. The advent of next-generation sequencing technology presents an unprecedented opportunity to identify pathogens in hitherto idiopathic diseases. Here we present PathSeq, a highly scalable software tool that performs computational subtraction on high-throughput sequencing data to identify non-human nucleic acids that may indicate candidate microbes. PathSeq exhibits high sensitivity and specificity in its ability to discriminate human from non-human sequences using both simulated and experimental transcriptome and whole-genome sequencing data. PathSeq is implemented in a cloud-computing environment making it readily accessible by the scientific community.

Introduction

Previously our group and others have developed a computational approach to pathogen discovery, sequence-based computational subtraction [3-6]. This method is based on the premise that infected tissues contain both human and microbial nucleic acids and that novel pathogen-derived sequences can be detected after subtracting human sequences. This unbiased approach to pathogen discovery is an advance over targeted PCR or pan-microbial array methods because it requires no sequence information ab initio about the organism being sought. However, performing computational subtraction at any significant scale was initially cost-prohibitive as this method requires a large number of input sequences, given that any pathogen present is likely to have low nucleic acid representation relative to that of the human host.

The recent development of next-generation sequencing methods [7,8], however, has made computational subtraction-based pathogen discovery a viable option. For example, massively parallel pyrosequencing combined with computational subtraction has resulted in the discovery of novel viruses in human disease: Merkel cell polyomavirus in Merkel cell carcinoma [9] and a novel Old World arenavirus in a cluster of patients with fatal transplant-associated disease [10]. Indeed, the past few years have seen steep drops in price and increases in throughput for next-generation sequencing technologies, and these trends are expected to accelerate in the near future [7,8]. However, this advancement in technology brings with it new computational challenges. Analyzing sequence data using the computational subtraction method is computationally expensive relative to most other next-generation sequencing analyses because it requires subtractive alignments to several large reference databases using local alignment algorithms such as BLAST.

Here we present PathSeq, a comprehensive computational tool for the analysis of the non-host portion of resequencing data that is capable of detecting the presence of both known and novel pathogens as well as any resident microorganisms. PathSeq runs efficiently on sequence datasets of any size in a scalable and completely reproducible fashion because it is developed on a parallel computing architecture and is implemented in a cloud-computing environment. The PathSeq software package is available for public use in the form of a machine image for cloud computing, which can be launched and monitored using no more than a basic laptop computer. We believe that PathSeq opens the way for a new large-scale effort in pathogen discovery by any researcher with access to deep sequencing data from human tissue.

Results and Discussion

The PathSeq process begins with a subtractive phase in which input reads are subtracted by alignment to human reference sequences (**Fig. 1-1a**), and continues with an analytic phase in which the remaining reads are aligned to microbial reference sequences and assembled *de novo*. The input reads are first filtered to remove low quality, duplicate, and repetitive sequences. The initial subtractive alignments are performed using the rapid short read aligner MAQ [11] against five reference human sequence databases, including both genomic DNA and transcriptome references (see **Material and Methods**). At the end of each subtractive alignment step, mapped reads are discarded and unmapped reads are subjected to further subtractive analyses. In the final steps, the residual reads are aligned to two additional human reference databases first using the Mega BLAST algorithm and then BLASTN. This identifies alignable reads with additional mismatches and/or short gaps that are not aligned by MAQ. The set of reads which remain unmapped after the subtractive phase are candidate non-human, pathogen-derived reads. A similar schema may be used for other host organisms by substituting the appropriate reference genome databases.

The analytic phase of PathSeq is composed of several steps that are performed in parallel (**Fig. 1-1b**). To identify previously sequenced microbes, all unmapped reads are aligned to reference viral, bacterial, and fungal sequence databases by BLASTN and BLASTX. To assess the bacterial composition of a sample containing a rich microbiome, PathSeq performs a metagenomic analysis by aligning all unmapped reads to the complete collection of currently sequenced whole bacterial genomes and quantifying bacterial representation by a measure of both the total number of aligned reads and the bacterial genome coverage (see **Materials and**

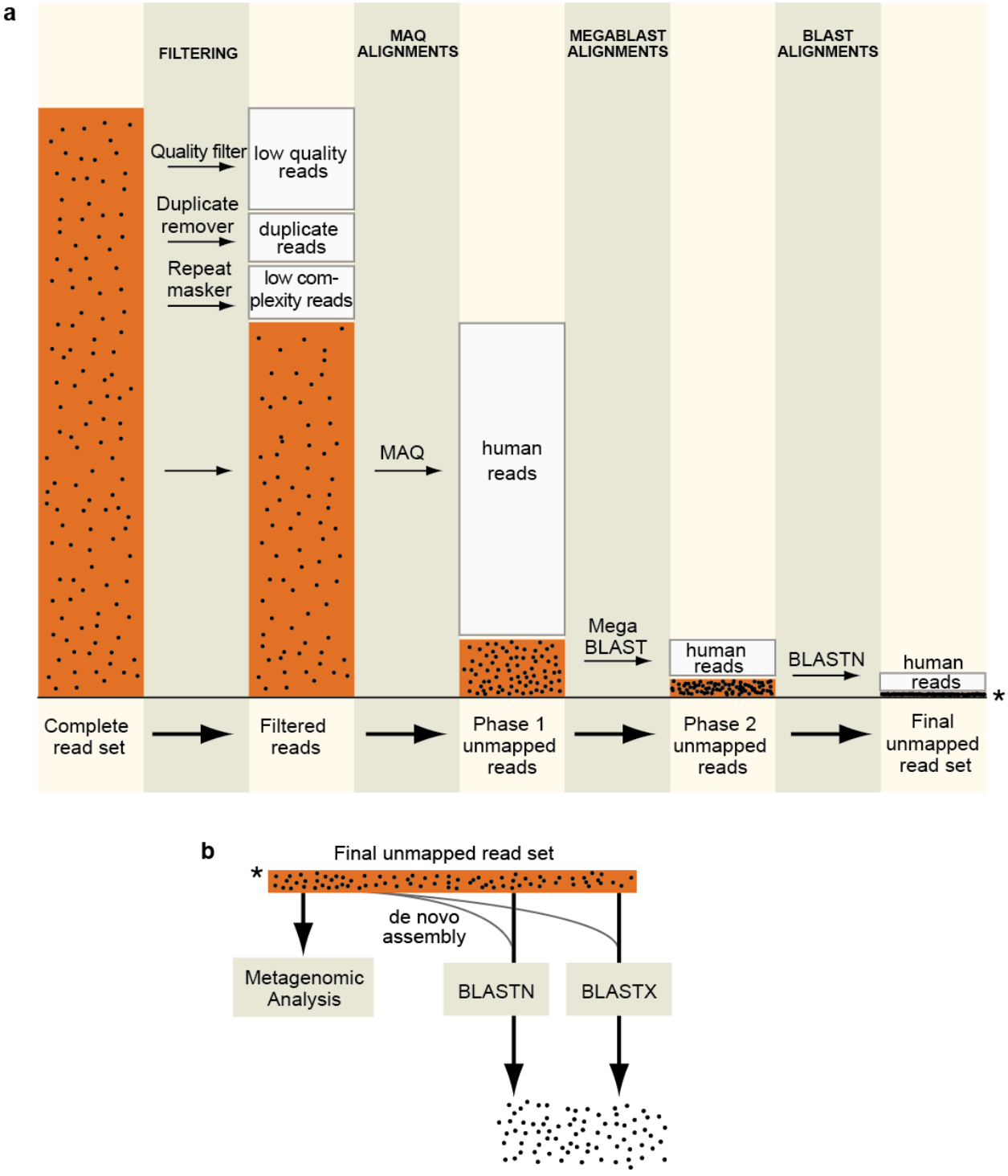


Figure 1-1. The PathSeq workflow. (a) Conceptual workflow of the subtractive phase of PathSeq. The size of the read set (orange bars) is proportional to the number of reads at the indicated step in a typical run of the method. The black dots in the bars represent pathogen-derived sequences which become progressively concentrated. The steps in this conceptual workflow have been reordered for concision (see **Materials and Methods** for actual ordering). **(b)** Conceptual workflow of the analytic phase of PathSeq. The asterisk indicates the unmapped readset that is carried over from the subtractive phase.

Methods, Supplementary Tables 1-1 and **1-2**). To increase the likelihood of discovering a novel organism, all unmapped reads are *de novo* assembled using the short read assembler Velvet [12]. The formation of large contigs composed of several unmapped reads that do not possess significant alignment similarity to any sequence in the reference databases may be suggestive of a previously undetected organism.

To demonstrate the utility of PathSeq, we used simulated data to assess the ability of the method to (i) efficiently subtract human-derived sequences and (ii) minimize the subtraction of microbe-derived sequences (**Fig. 1-2a**). We created a simulated sequence dataset by combining sequences generated from a reference human transcriptome database and several virus genomes (**Supplementary Fig. 1-1**). Twenty million 100-mers were randomly generated from the reference transcriptome. The simulated virus reads were generated from twelve viral genomes; each viral genome was substitutionally mutated randomly at twelve distinct rates (0, 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, and 90 percent), to simulate unknown viruses at different evolutionary distances from known viruses, producing 1,000 reads per mutated genome for a total of (12 x 12 x 1,000) 144,000 virus reads.

After the subtractive steps of the PathSeq pipeline, all 20 million human transcriptome-derived reads were correctly identified as human, and only 1,122 (0.78%) virus genome-derived reads were subtracted (**Fig. 1-2a**). Of these 1,122 reads, 1,120 were identified as repetitive sequences and the remaining 2 reads were subtracted because of alignment similarity to the human genome (**Supplementary Table 1-3**).

To model the performance of PathSeq on low-quality sequence data, we introduced "sequencing errors" into this same readset based on the distribution of actual Illumina sequencing errors and found that this did not significantly affect the performance of PathSeq

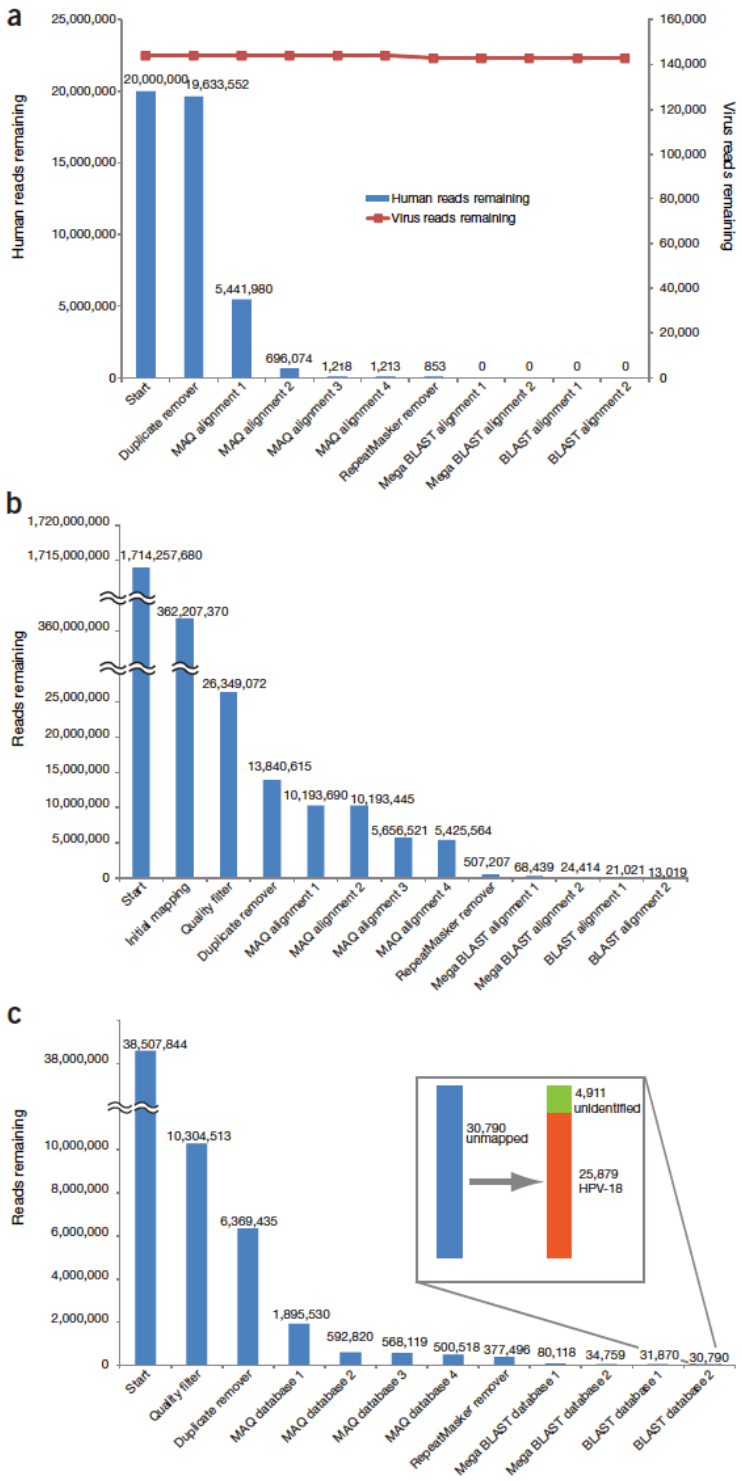


Figure 1-2. PathSeq performance on simulated and experimental sequence data. (a) Reads were generated by sampling random 100-mer sequences from a human transcriptome database to produce 20 million reads, and from a set of twelve virus genomes each substitutionally mutated at twelve distinct rates, generating 144,000 reads (see **Supplementary Fig. 1-1**). The blue bars represent the number of human reads remaining after the indicated step in the PathSeq workflow, and the red squares connected by a line represent the remaining viral reads. **(b)** We applied whole-genome sequencing data from a human ovarian tumor and **(c)** one lane of total-RNA transcriptome sequencing from HeLa cell lines to PathSeq. The inset in panel c shows that the 30,790 reads remaining after the subtractive phase of PathSeq are predominantly composed of HPV-18 sequences.

(**Supplementary Fig. 1-2, Supplementary Table 1-4, Materials and Methods**). With the exception of human rhinovirus A, which contained many repetitive sequences, greater than 97% of all non-mutated, virus-derived reads were correctly identified by alignment to the viral nucleotide database, and over 50% were still correctly identified at a substitutional mutation rate of over 20% (**Supplementary Fig. 1-3a**). Sequence alignment following *de novo* assembly allowed the identification of sequences with an even higher mutation rate (**Supplementary Fig. 1-3b**). We note that the presence of large, unidentifiable contigs in experimental sequence data could suggest the presence of a novel microbe lacking sequence homology to known microbes, and propose that such a result should justify follow-up by PCR, 3'- or 5'-RACE, and Sanger sequencing.

This notion of identifying the presence of microbes by contig formation prompted us to ask how many reads are required to form sufficiently large contigs. The probability of forming contigs from reads originating from a single genome is a function of two variables: (i) the size in base-pairs of the genome in question; and (ii) the number of reads derived from the genome. We simulated the ability of Velvet, the short read assembler that is used in PathSeq, to form contigs that are at least 1.75 times the size of the input reads from a genome by randomly generating reads from genomes of varying length (**Supplementary Fig. 1-4**). We found that there is a >75% chance of forming contigs from genomes as large as 20kb when only 20 reads are derived from the genome (using 100bp reads). This suggests that relying on contig assembly to indicate the presence of a novel genome may be a practical approach.

We then tested the performance of PathSeq on a set of sequences representing many-fold coverage of the whole genome of a serous ovarian carcinoma tumor that was sequenced as part of The Cancer Genome Atlas (see **Material and Methods**). Starting with slightly more than 1.7

billion reads of 101 bp each, we first removed mapped reads following initial alignment to the reference human genome, and then performed stringent quality filtering to yield 26.3 million reads. We ran these sequences on PathSeq and were left with a final set of 13,019 reads after subtraction, less than 0.001% of the original reads (**Fig. 1-2b**). The analytic phase of PathSeq did not yield any evidence that the remaining reads were derived from a pathogen; rather they likely represented yet-uncharacterized regions of the human genome or sequencing artifacts. This substantial subtraction efficiency, 99.9992%, demonstrates that the performance of PathSeq on simulated data can indeed be extended to real, human whole-genome sequencing.

We next generated sequence data from HeLa cervical cancer cell lines with the expectation of finding human papillomavirus (HPV) type18. We sequenced a cDNA library generated from total RNA isolated from HeLa cells on a single lane of Illumina sequencing, generating 10.3 million quality- and purity-filtered 76-bp reads. We applied this sequence data to PathSeq (**Fig. 1-2c**). Human-derived reads were efficiently subtracted during the subtractive phase such that 0.30% of input reads remained unmapped. Out of these 30,790 reads, 25,879 were identified as HPV-18, leaving 4,911 non-human, non-HPV-18 reads. We then collected all of the HPV-18 reads and searched for those whose pair-mate aligned to the human reference genome. This allowed us to identify the integration site of the viral genome in a region of chromosome 8q24 between positions 128,300,300 and 128,310,400 just upstream of the *MYC* oncogene.

One longstanding goal of computational subtraction is the identification and characterization of *every* read in a dataset. Although the sensitivity and specificity data for PathSeq are impressive, it still leaves 0.00076% of the ovarian whole-genome sequence reads and 0.013% of the HeLa RNA-Seq reads unaccounted for. This shortfall might be explained by

error-ridden sequences passing the quality filter, reads that map to the splice junction of rare splice variants in RNA-Seq data, or reads that map to regions of the human genome that have not yet been characterized. A number of groups have recently reported novel human genome sequences by *de novo* assembly of next-generation whole-genome sequence data [13,14] or by using a fosmid end-sequence pair mapping approach [15]. Ideally, these new sequences could form a subtractive database for PathSeq and help reduce the total number of unaccounted reads. However, by performing a simple Mega BLAST alignment of these sequences to microbial databases we find that many sequences from all three above-mentioned studies have perfect matches to known bacteria, fungi, and viruses, raising the question of whether they may contain novel microbes as well. Therefore, an essential next step in the improvement of computational subtraction-based methods is the creation of a reliable database of human genome scaffolds that extend the current human reference genome.

Taken together, our results demonstrate the ability of PathSeq to identify both known and novel microorganisms in high throughput human resequencing data. Just as current metagenomic surveys of the world's oceans and soils are yielding remarkable new organisms, so too do we expect to reveal new viruses, bacteria, and fungi in human tissue with important medical implications. We are making PathSeq available for public use at http://www.broadinstitute.org/software/pathseq/, and it is our hope that investigators will use this tool to join our efforts in pathogen discovery.

Materials and Methods

RNA-seq library construction from HeLa cells and sequencing

RNA was extracted from cultured HeLa cells according to the RNeasy Kit (Qiagen) protocol. cDNA sequencing library construction was performed as described previously [16], with noted modifications below. The cDNA library was sequenced on the Illumina Genome Analyzer II (GAII) platform. The mean fragment length was approximately 350 base pairs. One lane of paired-end, 76 base pair sequencing was performed, producing 38.5 million purity filtered reads, which yielded 10,304,513 high quality reads following quality filtering.

Modifications made to the Illumina RNA-Seq protocol

Total RNA (500 ng) was heated at 98ºC for 100 min in THE RNA Storage Solution (1 mM sodium citrate, pH 6.4; Ambion/ABI, AM7000) to fragment the RNA to a mean size of ~500 nucleotides. Quality of RNA fragmentations was assessed on a Bioanalyzer 2100 (Agilent). First-stand cDNA synthesis was performed by adding random hexamers (Invitrogen, 48190-011) to the RNA and heating at 70ºC for 10 min, and then immediately incubating at 50 ºC for 1 h upon addition of Superscript III reverse transcriptase (Invitrogen). Second-strand synthesis was carried-out with *E. coli* DNA ligase and *E. coli* DNA polymerase I (Invitrogen) for 2.5 h at 16 ºC. cDNA was purified using the MiniElute PCR Purification Kit (Qiagen) and evaluated using Bioanalyzer. End-repair, addition of adenine to the 3' end of the DNA fragments, and adapter ligation was performed as described in Guttman *et al.*, except a 2:1 molar ratio of adapter to DNA fragment was used during adapter-ligation. The resulting adapter-ligated fragments were purified on a 4% SeaKem LE agarose gel (Lonza) and a 400-500 base pair band was cut out of the gel and purified using the MiniElute kit. PCR was performed with Phusion DNA polymerase (Finnzymes) and adapter-specific primers using the following conditions: 2 min at 98 ºC; [10 s at 98 ºC, 30 s at 65 ºC, 30 s at 72 ºC] for 13 cycles; 5 min at 72 ºC. Following PCR, a second round

of gel extraction was performed as described above, and the product was submitted for Illumina sequencing.

The PathSeq workflow and cloud implementation

The PathSeq pipeline is designed using the Apache Hadoop implementation of the MapReduce programming framework (http://hadoop.apache.org/mapreduce) and can be run on the Amazon Elastic Compute Cloud (EC2) (http://aws.amazon.com/ec2/) [17]. The workflow is comprised of three modules: pre-subtraction, subtraction, and post-subtraction. The pre-subtraction module is simply a quality filtering step and is run on the user's local machine, whereas the subtraction and post-subtraction modules are executed on a Hadoop-based cluster (19 worker nodes and 1 master node) built using the Amazon Elastic Compute Cloud (Amazon EC2).

Amazon's Simple Storage Service (S3) file system (http://aws.amazon.com/s3/) is used to store the reference sequences and readset, and the config files and scripts are distributed across all nodes on the cluster using the Hadoop Distributed File System (HDFS). The reference sequences are continuously updated on the PathSeq system and users are given the option of substituting any built-in database with a database of their choice; however we provide data download dates for reference sequences used in experiments reported in this paper below.

All processes are run on the Hadoop cluster in multiple map phases. The subtraction module comprises of two mappers. First, subtractive alignments are performed with MAQ (Release 0.5.0, default settings) against a set of six human sequence databases: the 1000 Genomes Project female reference (ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/reference/, downloaded 2009-04-11), the Ensembl

Homo sapiens cDNA database (ftp://ftp.ensembl.org/pub/current/fasta/homo_sapiens/cdna/, downloaded 2009-04-22), the human genome and transcriptome BLAST database (ftp://ftp.ncbi.nih.gov/blast/db/, downloaded 2009-05-25), and the set of three assembled human genomes available on NCBI (hs_alt_Celera, hs_alt_HuRef, hs_ref_GRCh37, ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/Assembled_chromosomes/, downloaded 2009-06-19). The next map phase is composed of three steps: RepeatMasker (http://www.repeatmasker.org/), MegaBlast, and BLASTN. First, the reads are applied to RepeatMasker (version open-3.2.8, libraries dated 2009-06-04), and any reads with three or more masked nucleotides are discarded. Subtractive alignments are next performed using MegaBlast (Blast Tools version 2.2.23, cut-off expect value 10^{-7}, word size 16) to two human sequence databases: the NCBI *Homo sapiens* RNA database (ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/RNA/, downloaded 2009-11-20), and the Ensembl human genome reference (ftp://ftp.ensembl.org/pub/current/fasta/homo_sapiens/dna/, downloaded 2009-10-22). The final set of subtractive alignments are then performed with BLASTN (Blast Tools version 2.2.23, cut-off expect value 10-7, word size 7, nucleotide match reward 1, nucleotide mismatch reward -3, gap open cost 5, gap extension cost 2) to the same two databases. A reduce phase gathers all remaining reads into one consolidated file which serves as input to the post-subtraction module.

The post-subtraction module is also comprised of two mappers. The first mapper is a set of BLASTN (parameters as above) and BLASTX alignments (Blast Tools version 2.2.23, cut-off expect value 10-4, word size 3, matrix: BLOSUM62, gap open cost 11, gap extension cost 1) to viral (downloaded from NCBI Nucleotide (http://www.ncbi.nlm.nih.gov/nucleotide) using search term "'viruses'[porgn:__txid10239]" on 2010-02-26), fungal (downloaded using the term "'fungi'[porgn:__txid4751]" on 2009-11-23), bacterial and archaeal

(ftp://ftp.ncbi.nih.gov/genomes/Bacteria/, downloaded 2010-03-30), and non-redundant protein (ftp://ftp.ncbi.nih.gov/blast/db/nr, downloaded 2010-04-05) reference sequences. This mapper also performs a *de novo* assembly (Velvet 0.7.31, k-mer size 21) on the full set of reads remaining from the previous map phase. The complete post-subtraction BLAST output files as well as the full set of unmapped reads and contigs are then uploaded and stored on the S3 storage system.

The Amazon Machine Image (AMI) required to build the PathSeq Hadoop cluster is accessible from Amazon Web Services (http://www.broadinstitute.org/software/pathseq/).Pathseq is implemented in Python, Java, C++ and C shell, and has been tested on a Linux 2.6.18-194.8.1.e15 X86_64 system.

PathSeq runtime and performance

PathSeq analysis was performed using a cluster of 19 worker nodes and 1 master node, which were EC2 Large CPU instances (7GB of memory and 2 processor cores). Full analysis of HeLa cell RNA-Seq data described in this report was performed in approximately 13 hours (wall clock time) for a total price of $89 USD. The CPU time for this analysis was approximately 270 hours. Actual runtime and cost may vary depending on congestion on the Amazon EC2, Internet traffic, and the method of data upload. Because of its parallel architecture, PathSeq can analyze substantially larger datasets in a similar timeframe simply by increasing the cluster size.

Metagenomic analysis

The metagenomic analysis module of PathSeq reports the relative abundance of bacteria and archaea. This analysis begins with a MegaBlast alignment of the readset against the

complete set of fully sequenced bacterial and archaeal genomes (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/, downloaded 2010-03-30), reporting all hits with >90% sequence identity and >90% query coverage. The top 30 hits for each read are reported. Using these alignment results, classifications of each read are attempted at the phylum, then genus, then species level. If a given read cannot be classified uniquely at a given classification level (i.e. it has multiple hits to different reference sequences with equivalent E-values), then it is considered ambiguous and discarded from analysis at that level. Using species-level classifications, the fraction-genome-coverage is calculated for each species that received a hit, and this metric is used to quantify the relative abundance of a given species, normalized by the genome size.

Generation and analysis of simulated sequencing data

Simulated human transcriptome and virus sequence data. Twenty million 100-mers were randomly generated from a reference human transcriptome (ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/RNA/, downloaded 2009-11-20). Simulated virus reads were derived from twelve virus sequences: NCBI Nucleotide accession AY740741, CY000455, EU643590, FJ356716, FJ464337, GQ290462, GQ415051, NC_000883, NC_001405, NC_001806, NC_005179, and NC_007815. For each of these sequences, substitutional mutations were introduced at a frequency of 0%, 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%. In this process, nucleotides along the sequence are chosen at random with replacement (i.e. the same nucleotide can be chosen twice at random) and converted to a different nucleotide (for example, A is converted to C, G, or T). This produces 144 sequences (12 input sequences, each mutated at 12 frequencies). For each of these 144 sequences, 1,000 “reads” of length 100bp were produced at random. The resulting 144,000

simulated reads were pooled with the 20 million simulated human reads and analyzed on PathSeq.

Contig formation simulations. In this experiment, "genomes" of size 200bp to 20,000bp in increments of 200bp were generated from the Human herpesvirus 5 genome (accession GQ466044). For each of these 100 "genomes", two to twenty 100-mer sequences ("reads") were generated at random. This process was performed eleven times. For each "genome" size, "read" number pair, Velvet assembly with k-mer size 21 was performed. The frequency with which contigs of at least 175bp were generated was recorded.

Introduction of "sequencing errors" into simulated sequence data

Sequencing errors were introduced into the simulated reads based on quality scores seen in a whole-genome sequence dataset of a glioblastoma multiforme (GBM) primary tumor (sequenced as part of The Cancer Genome Atlas; data available via the NCBI Short Read Archive (SRA) identifier SRX010704). The average quality score for each base along the length of the reads was calculated across the dataset and offset by -5 (**Supplementary Fig. 1-3a**). This was converted into a probability value and used to "mutate" our simulated reads (i.e. for a sequence error probability of 0.001, there is a 0.1% chance that the base will be converted to a different base).

Human whole-genome ovarian tumor sequencing data

The human ovarian tumor whole-genome sequencing dataset was sequenced as part of The Cancer Genome Atlas, and the data is available via the NCBI SRA identifier SRX010747.

This is a 101 base pair, paired-end sequence dataset with a nominal fragment length of 264 base pairs.

References

1. Relman DA: **The search for unrecognized pathogens.** *Science* 1999, **284**:1308–1310.

2. Lipkin WI: **Microbe hunting.** *Microbiol. Mol. Biol. Rev.* 2010, **74**:363–377.

3. Weber G, Shendure J, Tanenbaum DM, Church GM, Meyerson M: **Identification of foreign gene sequences by transcript filtering against the human genome.** *Nat. Genet.* 2002, **30**:141–142.

4. Xu Y, Stange-Thomann N, Weber G, Bo R, Dodge S, David RG, Foley K, Beheshti J, Harris NL, Birren B, et al.: **Pathogen discovery from human tissue by sequence-based computational subtraction.** *Genomics* 2003, **81**:329–335.

5. Tengs T, LaFramboise T, Den RB, Hayes DN, Zhang J, DebRoy S, Gentleman RC, O'Neill K, Birren B, Meyerson M: **Genomic representations using concatenates of Type IIB restriction endonuclease digestion fragments.** *Nucleic Acids Res* 2004, **32**:e121.

6. Feng H, Taylor JL, Benos PV, Newton R, Waddell K, Lucas SB, Chang Y, Moore PS: **Human transcriptome subtraction by using short sequence tags to search for tumor viruses in conjunctival carcinoma.** *The Journal of Virology* 2007, **81**:11332–11340.

7. Mardis ER: **Next-Generation DNA Sequencing Methods**. *Annual Review of Genomics and Human Genetics* 2008, [no volume].

8. Shendure J, Ji H: **Next-generation DNA sequencing**. *Nat. Biotechnol.* 2008, **26**:1135–1145.

9. Feng H, Shuda M, Chang Y, Moore PS: **Clonal integration of a polyomavirus in human Merkel cell carcinoma.** *Science* 2008, **319**:1096–1100.

10. Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, Conlan S, Quan P-L, Hui J, Marshall J, et al.: **A new arenavirus in a cluster of fatal transplant-associated diseases.** *N. Engl. J. Med.* 2008, **358**:991–998.

11. Li H, Ruan J, Durbin R: **Mapping short DNA sequencing reads and calling variants using mapping quality scores.** *Genome Res* 2008, **18**:1851–1858.

12. Zerbino DR, Birney E: **Velvet: algorithms for de novo short read assembly using de Bruijn graphs.** *Genome Res* 2008, **18**:821–829.

13. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen Y-J, Makhijani V, Roth GT, et al.: **The complete genome of an individual by massively parallel DNA sequencing.** *Nature* 2008, **452**:872–876.

14. Li R, Li Y, Zheng H, Luo R, Zhu H, Li Q, Qian W, Ren Y, Tian G, Li J, et al.: **Building the sequence map of the human pan-genome.** *Nat. Biotechnol.* 2010, **28**:57–63.

15. Kidd JM, Sampas N, Antonacci F, Graves T, Fulton RS, Hayden HS, Alkan C, Malig M, Ventura M, Giannuzzi G, et al.: **Characterization of missing human genome sequences and copy-number polymorphic insertions**. *Nat. Methods* 2010, **7**:365–U47.

16. Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, et al.: **Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs.** *Nat. Biotechnol.* 2010, **28**:503–510.

17. Wall DP, Kudtarkar P, Fusaro VA, Pivovarov R, Patil P, Tonellato PJ: **Cloud computing for comparative genomics.** *BMC Bioinformatics* 2010, **11**:259.

CHAPTER 2

Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma

This Chapter is a reproduction of a published manuscript:

Kostic, A.D., Gevers, D., Pedamallu, C.S., Michaud, M., Duke, F., Earl, A.M., Ojesina, A.I., Jung, J., Bass, A.J., Tabernero, J., Baselga, J., Liu, C., Shivdasani, R.A., Ogino, S., Birren, B.W., Huttenhower, C., Garrett, W.S., and Meyerson, M. (2012). Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. Genome Res 22, 292–298.

Author Contributions:

A.D.K. and M.M. conceived the study. A.D.K., D.G., C.S.P. and C.H. performed all computational analyses. A.D.K., M. Michaud and F.D. performed bench experiments. A.J.B., J.T., J.B., C.L., R.A.S. and S.O. provided samples for analysis. A.D.K., D.G., C.S.P., A.M.E., A.I.O., J.J., A.J.B., B.W.B., C.H., W.S.G. and M.M. contributed to experimental design and data interpretation. A.D.K., W.S.G. and M.M. wrote the manuscript.

Abstract

The tumor microenvironment of colorectal carcinoma is a complex community of genomically altered cancer cells, non-neoplastic cells, and a diverse collection of microorganisms. Each of these components may contribute to carcinogenesis, however the role of the microbiota is the least well understood. We have characterized the composition of the microbiota in colorectal carcinoma using whole genome sequences from 9 tumor/normal pairs. *Fusobacterium* sequences were enriched in carcinomas, confirmed by quantitative PCR and 16S rDNA sequence analysis of 95 carcinoma/normal DNA pairs, while the Bacteroidetes and Firmicutes phyla were depleted in tumors. Fusobacteria were also visualized within colorectal tumors using FISH. These findings reveal alterations in the colorectal cancer microbiota; however, the precise role of fusobacteria in colorectal carcinoma pathogenesis requires further investigation.

Introduction

Malignant tumors are complex communities of oncogenically transformed cells with aberrant genomes, associated non-neoplastic cells including immune and stromal cells, and sometimes microbes, including bacteria and viruses. Several viruses that can integrate into the human genome directly cause cancer, such as human papillomavirus in cervical cancer [1] and Kaposi's sarcoma-associated herpesvirus in Kaposi's sarcoma [2]. In other cases, microorganisms lead indirectly to cancer through chronic inflammatory responses—a mechanism by which *Helicobacter pylori* contributes to both gastric cancer and MALT lymphoma [3,4].

In the human distal gut, where microbial cells outnumber host cells nine-to-one [5], the microbiome can impart both beneficial and detrimental effects on host physiology contributing to health or disease susceptibility. Gut microbial communities (microbiota) may also influence the development of colorectal carcinoma [6-8]. Sears and Pardoll have recently introduced the concept of the "alpha-bug" – wherein select members of a microbial community, in addition to possessing virulence and pro-carcinogenic features, are capable of remodeling the microbiome as a whole to drive pro-inflammatory immune responses and colonic epithelial cell transformation leading to cancer [9].

We postulate that if the microbiota play an active role in the pathogenesis of colorectal carcinoma, then these microbes will be found within the tumor microenvironment, and the composition of the tumor microbiome will differ from that of adjacent non-neoplastic tissue. We have undertaken unbiased, sequence-based approaches, followed by cytological analysis, to probe the differences in the microbial composition of the colorectal carcinoma tumor microenvironment relative to adjacent non-neoplastic tissue. We now report an association of *Fusobacterium* with the colonic mucosa of colorectal carcinoma.

Results

To determine the microbial composition of human colorectal cancer, we analyzed whole genome sequences of nine colorectal cancers and matched normal colons [10] using PathSeq, a computational subtraction pipeline that culls out candidate microbial sequences [11]. These presumed bacterial sequences were identified by alignment to known sequenced microbial genomes (**Fig. 2-1a**, **Supplementary Fig. 2-1**, **Supplementary Table 2-1**). PathSeq analysis

also identified the presence of virus sequences in these specimens including human herpesvirus 7, however no significant differences in viral sequence levels were detected between tumor and normal DNA (**Supplementary Table 2-2**). Hierarchical clustering analysis of the species-specific relative abundances of microbial sequences revealed that the microbial communities of a tumor and matched non-cancerous colon from a given patient were more similar to each other than are tumors or non-affected colon samples from different patients (**Fig. 2-1b**). This finding suggests that a patient's intestinal ecosystem may be more significant in shaping the microbiota than the generic microenvironment of a colon tumor or normal colonic tissue.

To identify bacterial species whose sequences are more abundant in colorectal tumors than in the matched, non-cancerous colorectal tissue, we applied a metagenomic biomarker discovery approach, LEfSe (Linear Discriminant Analysis (LDA) coupled with effect size measurements), which performs a non-parametric Wilcoxon sum-rank test followed by LDA analysis to assess the effect size of each differentially abundant taxon [12]. Using LEfSe, we found that *Fusobacterium* sequences were significantly enriched in the colorectal cancer metagenomes as were sequences from the family of *Streptococcaceae* (**Fig. 2-1c, Fig. 2-1d, Supplementary Fig. 2-2**).

As our initial screen was performed on a sample size of 9 cases, we next examined a larger cohort of 95 paired specimens of colon cancer and normal colonic DNA to survey the colon cancer microbiome and validate the tumor-specific enrichment of *Fusobacterium*. We amplified ribosomal 16S rDNA by PCR using consensus primers from 95 tumor/normal pairs, followed by pyrosequencing to assess the relative abundance of DNA from bacterial species

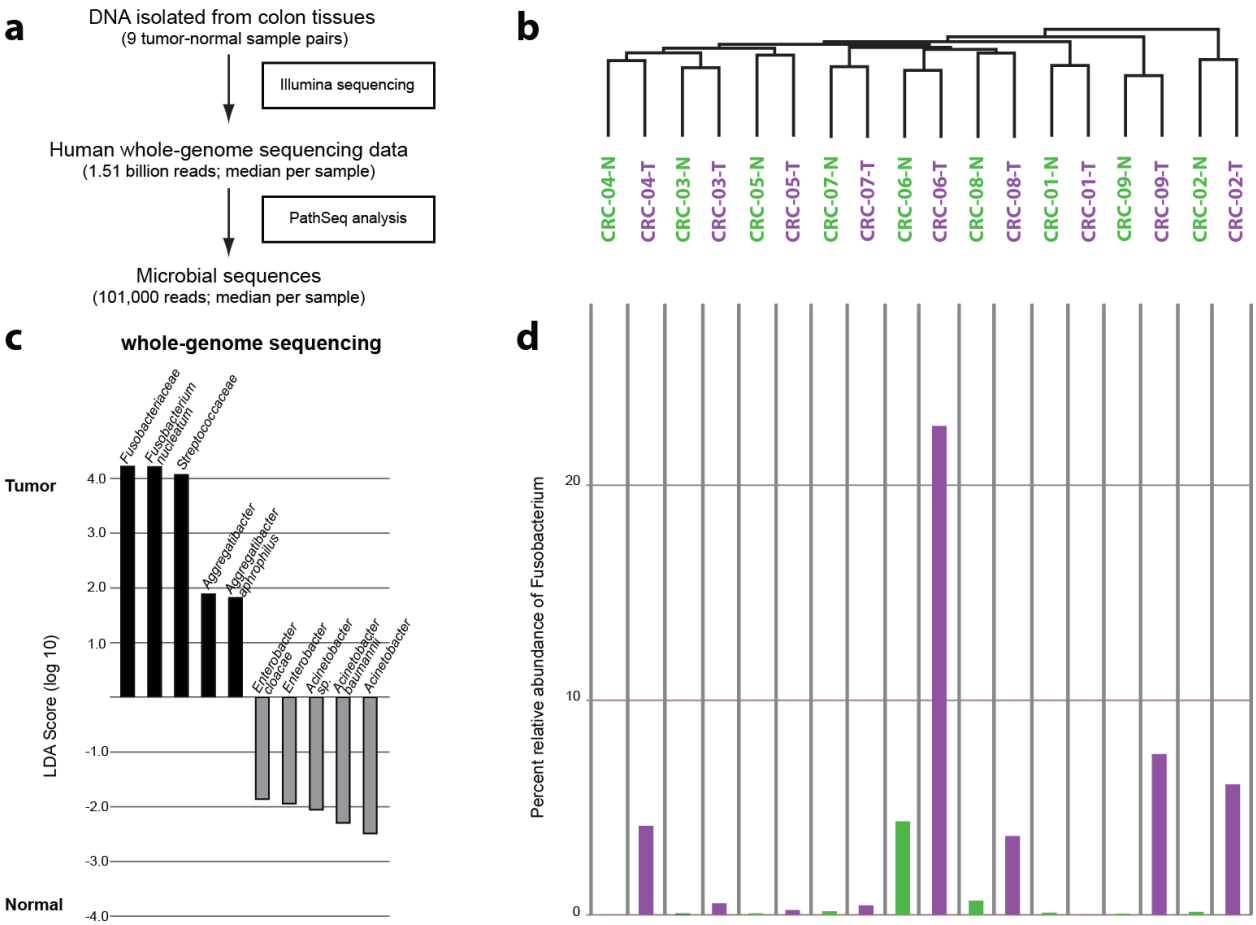


Figure 2-1. Whole-genome sequencing analysis of the colorectal cancer microbiome. (a) Schematic of experimental and computational whole-genome sequencing analysis workflow. **(b)** Hierarchical clustering of phylotype relative abundance measurements demonstrates that microbial composition of tumor/normal pairs within individuals is more highly correlated than tumor/tumor pairs or normal/normal pairs from different individuals. Normal samples are shown in green, tumors are shown in purple. **(c)** Linear Discriminant Analysis (LDA) coupled with effect size measurements identifies *Fusobacterium* as the most differentially abundant taxon in colon tumor versus normal specimens by whole-genome sequencing in 9 individuals. Tumor-enriched taxa are indicated with a positive LDA score (black), and taxa enriched in normal tissue have a negative score (gray). Only taxa meeting an LDA significant threshold of 1.8 are shown. **(d)** Percent relative abundance for the genus *Fusobacterium* is depicted across all samples in the order of the labels in (b), demonstrating a tumor-enrichment in most individuals.

(**Fig. 2-2a**). Overall, as was the case in our whole-genome sequence data, tumor/normal pairs from the same individual are much more highly correlated than tumor/tumor pairs or normal/normal pairs from different individuals (**Fig. 2-2b**). Colorectal tumors were associated with broad phylum-level changes including the depletion (i.e. reduced relative abundance) of

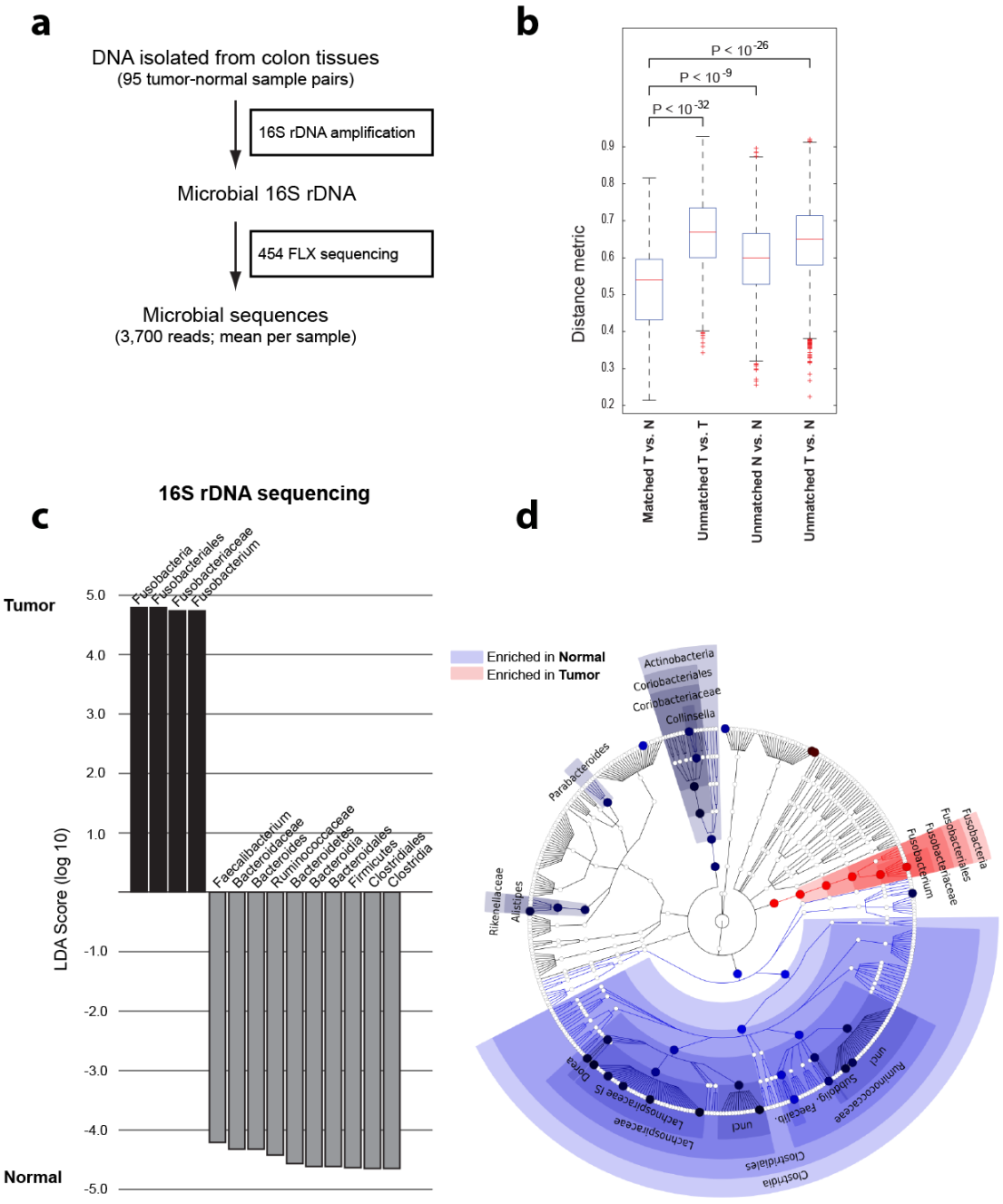


Figure 2-2. 16S rDNA sequencing analysis of the colorectal cancer microbiome. (a) Schematic of experimental and computational 16S rDNA sequencing analysis workflow. **(b)** Beta-diversity distances calculated using phylotype relative abundance measurements between all pairs of samples demonstrate that the microbial composition of tumor/normal pairs within individuals is more highly correlated than tumor/tumor pairs, normal/normal pairs, or tumor/normal pairs from different individuals. **(c)** Linear Discriminant Analysis (LDA) coupled with effect size measurements identifies *Fusobacterium* as the most differentially abundant taxon in colon tumor versus normal specimens by 16S rDNA sequencing in 95 individuals. Tumor-enriched taxa are indicated with a positive LDA score (black), and taxa enriched in normal tissue have a negative score (gray). Only taxa meeting an LDA significant threshold of 4.2 are shown. **(d)** A cladogram representation of data in (c). Tumor-enriched taxa are indicated in red, and taxa enriched in normal tissue are blue. The brightness of each dot is proportional to its effect size.

Firmicutes and Bacteroidetes, most prominently the Clostridia (**Fig. 2-2c, Fig. 2-2d, Supplementary Fig. 2-3**); however the overall diversity in the tumors relative to adjacent tissue was not significantly different (**Supplementary Fig. 2-4a** and **Supplementary Fig. 2-4b**). Consistent with our whole-genome sequencing results, the relative abundance of *Fusobacterium* was highly enriched in the population of tumor versus normal samples (**Fig. 2-2c** and **Fig. 2-2d**). However a tumor-enrichment for *Streptococcaceae* was not reproduced, most likely due to small sample-size in our initial whole-genome sequencing results. In addition, we analyzed patient metadata to identify correlations or possible confounding effects (including patient age, gender, ethnicity, tumor anatomic location, tumor purity, inflammation, necrosis, and vascularization), but only found a modest correlation with patient geographic location (**Supplementary Fig. 2-5**), as well as a correlation of higher microbial diversity with tumors of higher histological stage or grade (**Supplementary Fig. 2-4c**). The correlation of fusobacteria abundance with geographic location may either indicate a real geographic effect or else a confounder introduced by slightly differing sample collection protocols at the collection sites, for example time between surgery and freezing (see **Materials and Methods**).

As we have shown that *Fusobacterium* species are enriched in colorectal cancer DNA and tissue, we sought associations that might suggest that fusobacteria are required for the survival or maintenance of colorectal cancer cells. Because *Fusobacterium* species can invade colonic epithelial cells [13], we examined colorectal cancer cell lines and hepatic and lymph node metastases for evidence of fusobacterial DNA. Quantitative PCR analysis of 59 human colorectal cancer cell line DNAs revealed no significantly detectable *Fusobacterium* DNA, however these *in vitro* passaged cell lines are often cultured in the presence of antibiotics (**Supplementary Table 2-3**). Strikingly, however, when we examined surgically resected

colorectal cancer metastases, *Fusobacterium* was detected in 2 out of 11 cases (**Supplementary Table 2-4**).

Given the increased abundance of *Fusobacterium* sequences in colon cancer DNA, we next asked whether *Fusobacterium* could be detected in histological sections of colon cancer, and if so, where. To address this question, we used 16S rDNA fluorescence *in situ* hybridization (FISH) oligonucleotide probes on colonic biopsy sections. Employing probeBase consortium 16S rDNA probes that detect the majority of bacteria (EUB338) and members of the genus *Fusobacterium* (FUSO) [14,15], we performed FISH analysis on frozen (9 cases) and formalin-fixed paraffin embedded (12 cases) tissue sections from colorectal cancer and normal colon. The *Fusobacterium* probes detected bacteria in the colorectal cancer and normal tissue sections and were quantitated within the lamina propria and mucus (**Fig. 2-3a**); z-section stacks suggest that some of the imaged bacteria may reside intracellularly (data not shown). Consistent with the analysis of *Fusobacterium* DNA described above, FISH-detected fusobacteria were enriched in the colorectal cancer compared to the normal samples (**Fig. 2-3b, Supplementary Fig. 2-7**; see **Supplementary Fig. 2-8** for a comparison of *Fusobacterium* quantitation across all 4 methods), in contrast to total bacteria counts which were more evenly distributed (**Supplementary Fig. 2-7**).

Finally, we sought to assess the specific *Fusobacterium* species that are enriched in colorectal carcinomas. Based on the 16S ribosomal DNA sequences, 5 out of a total of 409 operational taxonomic units (OTUs, a proxy for species) identified in our samples were classified as members of the *Fusobacterium* genus. By performing multiple sequence alignments using our 5 OTUs along with 16S rDNA sequences from a reference set of 31 *Fusobacterium* species and constructing maximum likelihood trees, the OTUs were identified as most closely related to

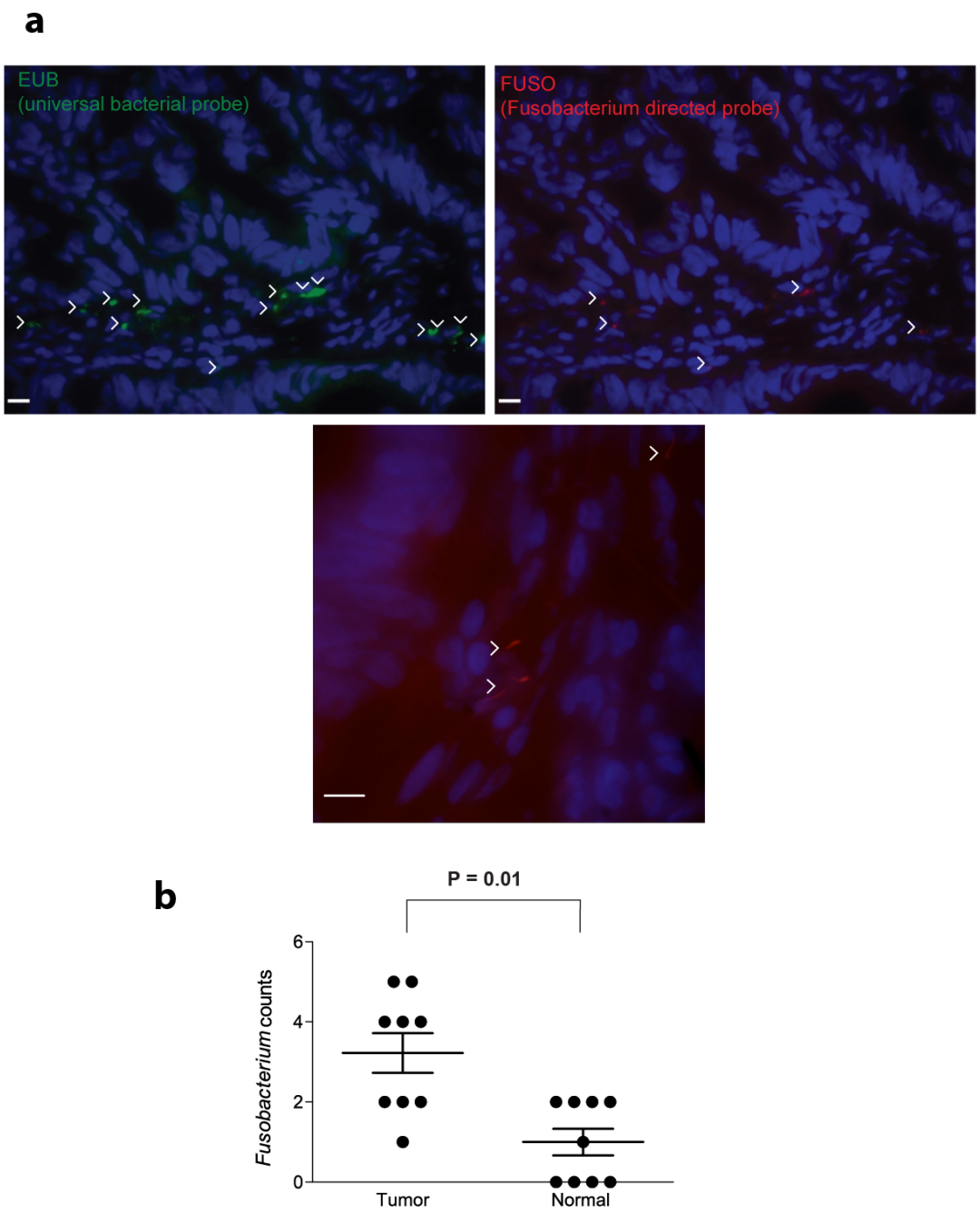


Figure 2-3. Fluorescence *in situ* hybridization (FISH) detects enrichment of fusobacteria in colorectal tumors. **(a)** FISH using an Oregon-Green 488-conjugated ‘‘universal bacterial’’ 16S rDNA-directed oligonucleotide probe (EUB338, green) (top left) and Cy3-conjugated *Fusobacterium* (FUSO, red) (top right and bottom center)16S rDNA-direct oligonucleotide probe demonstrates the presence of bacteria and *Fusobacterium* within in the colonic mucosa of colorectal tumor samples. Representative images are shown with a 10 μm scale bar in the lower corner of each panel; white arrowheads mark bacteria. Epithelial cell nuclei were stained with DAPI. **(b)** To determine whether *Fusobacterium* was enriched in tumor versus normal pairs, 3 random 40X fields were chosen for scoring by an observer blinded to tumor/normal status, using selection criteria of mucosal tissue depth and a minimum of 5 bacteria visualized by the EUB338 probe per field. Each dot represents data from either a tumor or normal sample from 9 tumor/normal paired cases. The mean, SEM, and p-values (calculated by a Wilcoxon matched-pairs signed rank test) are shown.

Fusobacterium nucleatum, *Fusobacterium necrophorum*, *Fusobacterium mortiferum*, and *Fusobacterium perfoetens* (**Fig. 2-4a**). The percent relative abundance in colorectal tumors versus normal colons of the two most abundant OTUs is shown in **Fig. 2-4b**, demonstrating that for most patients these OTUs are enriched in the tumor. Strikingly, only a subset of the cancers showed dramatic enrichment of *Fusobacterium* species, accounting for up to 89% of total bacterial DNA in some specimens; this result suggests that fusobacteria may be uniquely related to pathogenesis of subsets of colorectal cancer. The OTU with the greatest similarity to *F. nucleatum* was the most dominant phylotype identified within cancers, however some tumors contain more than one dominant species (**Supplementary Fig. 2-9**).

Discussion

In summary, genomic analysis of the microbiome of colorectal carcinomas reveals a significant enrichment of *Fusobacterium* species in these cancers, especially phylotypes with the greatest similarity to *F. nucleatum, F. mortiferum*, and *F. necrophorum.* This enrichment is confirmed by histological analysis of tumor tissue, and also the identification of *Fusobacterium* DNA in colon tumor metastases. Our analysis also reveals broader changes in the tumor environment such as the depletion of the Bacteroidetes and Firmicutes phyla, most notably the order Clostridiales. *Fusobacterium* species may have a fitness advantage in the evolving tumor microenvironment resulting in an altered microbiota in accordance with the "alpha-bug" hypothesis.

Interestingly, *Fusobacterium* species may be associated with inflammatory bowel diseases (IBD) including both ulcerative colitis and Crohn's disease

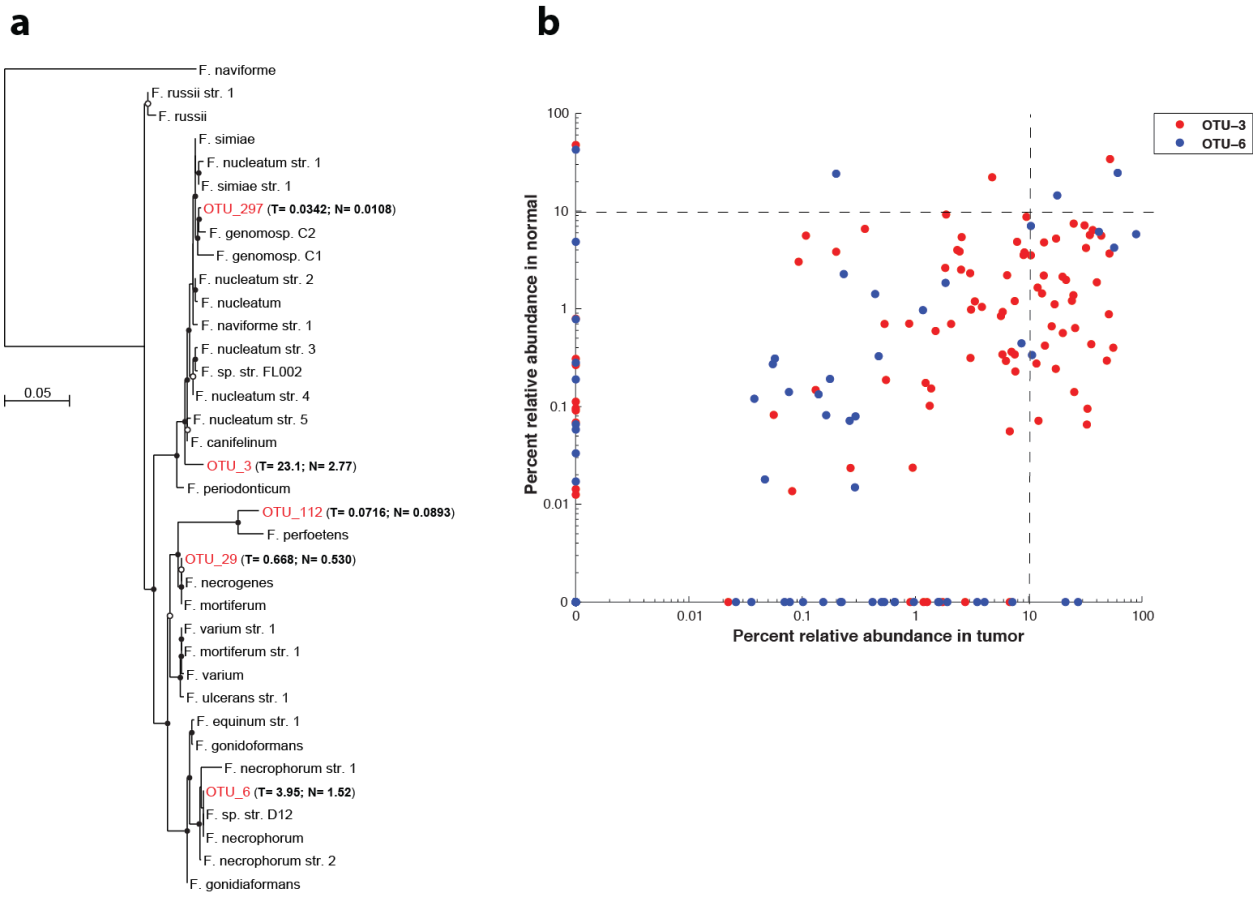


Figure 2-4. Phylogenetic analysis identifies several *Fusobacterium* species in human colon cancer tissues. (a) Approximately-maximum-likelihood phylogenetic trees were constructed on the V3-V5 region of the 16S rDNA gene using 31 reference *Fusobacterium* species along with the five most prominent OTUs identified in colon cancer specimens (indicated in red). Nodes that have bootstrap support above 50% and 75% are indicated with a white and black dot, respectively. The mean percent relative abundance in tumor (T) and normal (N) of each OTU is indicated in parentheses. The full names of the reference strains appear in **Supplementary Table 2-5**. **(b)** The abundance of the indicated OTU relative to all other phylotypes in a given specimen is shown for the two most abundant *Fusobacterium* OTUs in tumors (x-axis) and normal colon tissue (y-axis); each point represents tumor and normal abundance data for a different individual. The lower-right quadrant of the graph highlights the substantial proportion of patients for whom the *Fusobacterium* abundance is >10% in tumors but <10% in the matched normal.

[13,16,17], and IBD is a known risk factor, indeed one of three highest risk factors, for colorectal cancer. Furthermore, consistent with our findings in colorectal carcinoma, others have reported that several *Fusobacterium* strains were associated with IBD, however the majority (69%) were specifically associated with *F. nucleatum* [13]. Therefore it is worth further exploration of a causal link between *Fusobacterium* spp. with inflammatory bowel disease and colorectal carcinoma pathogenesis.

F. nucleatum and other *Fusobacterium* species can elicit host pro-inflammatory response [18] and posses virulence characteristics that promote their adhesiveness to host epithelial cells [19,20] and their ability to invade into epithelial cells [13,21]. Therefore, our findings of a tumoral enrichment of *Fusobacterium* spp. in colorectal carcinoma suggest the possibility that these organisms may contribute to tumorigenesis, perhaps in a limited subset of patients, most conceivably by an inflammatory-mediated mechanism. Alternatively, it is possible that fusobacteria accumulate in the tumor microenvironment in the late stages of tumorigenesis and therefore do not have a significant role in tumor development. Our results do not prove a causal relationship between *Fusobacterium* and colorectal cancer; the establishment or repudiation of such a relationship will require further studies of colorectal cancer in both human subjects and animal models of the disease. Additionally, case-control studies comparing tumor microbiota to that of colonic epithelial tissues from healthy individuals will serve to demonstrate whether *Fusobacterium* species are more prevalent in individuals with colon cancer relative to the general population.

In summary, our findings reveal species-specific alterations in the colorectal cancer microbiota, which may lead to microbiota-directed prevention, diagnostic, prognostic, and treatment strategies for these cancers.

Materials and Methods

Sample collection and preparation

Colorectal adenocarcinoma and adjacent non-affected tissue was obtained from the Vall d'Hebron University Hospital in Barcelona, Spain and Genomics Collaborative Inc. (GCI), using the sample collection protocols detailed below:

Vall d'Hebron University Hospital

All Frozen tissue samples were collected following the Standard Operating Procedures at the site. Participants were enrolled in the study prior to surgery, and informed consent was obtained by the surgeon. After arterial ligation and surgical removal of the tissues, the specimens were immediately transferred from the Operating Room to the Pathology suite and subsequently evaluated by the pathologist and, if possible, one fragment of healthy tissue and one of tumor were chosen and placed in a cryotube and frozen immediately in liquid nitrogen (time interval between specimen removal and freezing: maximum= 30 min.; median= 22 min.). The frozen tissue was stored at -80 °C until DNA extraction. All clinical data, captured on Case Report Forms, was double-data entered in to a clinical database.

Genomics Collaborative Inc.

All Frozen tissue samples were collected following the same Standard Operating Procedure at all collection sites, both within the US and in Vietnam. Participants were enrolled in to the study prior to surgery, and informed consent was obtained. After arterial ligation and surgical removal of the tissues, the tissue was transferred from the Operating Room to the Pathology suite. The time taken from surgical removal of the specimen until the time it was received for processing in the pathology suite, as well as the temperature at which it was transferred (room temp, or on ice) were recorded. The SOP required the Tissue Transmittal time (from removal to freezing of the sample) to ideally be between 30 to no more than 45 minutes. Upon dissection in the pathology suite, samples were cut into approximately 1g pieces and placed into pre-labed cyrovials

supplied in the sample kits. Samples were immediately frozen in liquid nitrogen vapor faze in charged vapor-shippers located at each site. An H&E slide was cut from the adjacent face of each tissue sample and sent to GCI along with the frozen tissue samples. Upon receipt at GCI, the samples were qualified by pathologist review of the H&E slide, and all samples were stored in Liquid N2 vapor freezers until requested for research. All clinical data, capture on Case Report Forms, was double-data entered in to a clinical database.

DNA extraction, whole genome sequencing, and analysis

DNA was extracted from colorectal carcinoma tumors and adjacent non-affected tissues and whole genome sequencing was performed as described previously [10]. Initial alignments to the human reference genome were performed as described [10]. All unaligned sequencing reads were (1) analyzed on PathSeq and (2) aligned to the complete set of fully sequenced bacterial and archaeal genomes (ftp://ftp.ncbi.nih.gov/genomes/Bacteria, downloaded 2010-10-07) by MegaBlast (Blast Tools version 2.2.23, word size 16, match reward 1, mismatch reward -2, gap open reward -5, gap extension reward -2). The top 30 sequence matches with >90% sequence identity and >90% query coverage were reported for each read (i.e. query). Classifications were performed at the domain, then phylum, then genus, then species level requiring unique alignments (i.e. reads with equivalent E-values to multiple taxa were removed from analysis). At the species level, relative abundance (RA) for each organism was calculated as follows: RA = (# unique alignment positions in genome * 1,000,000) / (# total alignable reads * genome size). The RA values were then per-sample normalized such that the total relative abundance for each sample sums to one. The resulting normalized RA matrix was analyzed on LEfSe [12].

Amplification and 454 sequencing of 16S gene

The 16S gene dataset consists of 454 FLX Titanium sequences spanning the V3 to V5 variable regions obtained for 190 samples (95 pairs). Detailed protocols used for 16S amplification and sequencing are available on the HMP Data Analysis and Coordination Center website (http://www.hmpdacc.org/tools_protocols/tools_protocols.php). In brief, genomic DNA was subjected to 16S amplifications using primers designed incorporating the FLX Titanium adapters and a sample barcode sequence, allowing directional sequencing covering variable regions V5 to partial V3 (Primers: 357F 5'-CCTACGGGAGGCAGCAG-3' and 926R 5' CCGTCAATTCMTTTRAGT-3'). Polymerase chain reaction (PCR) mixtures (25µl) contained 10ng of template, 1x Easy A reaction buffer (Stratagene, La Jolla, CA), 200mM of each dNTP (Stratagene), 200nM of each primer, and 1.25U Easy A cloning enzyme (Stratagene). The cycling conditions for the V3-V5 consisted of an initial denaturation of 95°C for 2 min, followed by 25 cycles of denaturation at 95°C for 40 sec, annealing at 50°C for 30 sec, extension at 72°C for 5 min and a final extension at 72°C for 7 min. Amplicons were confirmed on 1.2% Flash Gels (Lonza, Rockland, ME) and purified with AMPure XP DNA purification beads (Beckman Coulter, Danvers, MA) according to the manufacturer and eluted in 25 µL of 1X low TE buffer (pH 8.0). Amplicons were quantified on Agilent Bioanalyzer 2100 DNA 1000 chips (Agilent Technologies, Santa Clara, CA) and pooled in equimolar concentration. Emulsion PCR and sequencing were performed according to the manufacturer's specifications.

Processing of 16S sequence data

Resulting sequences were processed using a data curation pipeline implemented in mother [22], complimented by abundantOTU [23], and custom PERL scripts. Sequences were

removed from the analysis if they were <200 nt or >600 nt, had a read quality score <25, contained ambiguous characters, had a non-exact barcode match, or did show more than four mismatches to the reverse primer sequences (926R). Remaining sequences were assigned to samples based on barcode matches, after which barcode and primer sequences were trimmed and reads were oriented such that all sequences begin with the 5' end according to standard sense strand conventions. All sequences were aligned using a NAST-based sequence aligner to a custom reference based on the SILVA alignment [22,24]. Chimeric sequences were identified using the mothur implementation of the ChimeraSlayer algorithm [25]. Quality filtered and chimera-free sequences were clustered into Operational Taxonomic Units (OTU's) using abundantOTU [23]. Representative sequences per OTU were classified with the MSU RDP classifier v2.2 [26], maintained at the Ribosomal Database Project (RDP 10 database, version 6).

Quantitative PCR analysis

Quantitative real-time PCR was performed as described [27] using pan-*Fusobacterium* probe-primer sets as described [27]. *Fusobacterium* quantitation was measured relative to human endogenous 18S (Applied Biosystems TaqMan® Ribosomal RNA Control Reagents, Hs99999901_s1 (part number 4331182)).

Microbial FISH analysis

Frozen sections were fixed in Carnoy's solution overnight and embedded in paraffin, and 5 mm thick sections prepared and hybridized as previously described [15]. The sequences of the following FISH probes were obtained from probeBase (http://www.microbial-ecology.

net/probebase/) [14]: the ‘‘universal’’ bacterial probe-EUB338 (pB-00159), Fusobacterium targeted probe (pB-00782). Slides were imaged on an Olympus B40 microscope, digitally photographed using IP Lab. Three random fields per sample were chosen by an obsever blinded to tumor/normal status, using selection criteria of mucosal tissue depth and a minimum of 5 bacteria visualized by the EUB338 probe per field. Composite z-stacks were assembled in IP Lab and composite photomicrographs were assembled in Adobe Photoshop.

Data access

The 16S sequence data from this study have been submitted to the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/Traces/ sra/sra.cgi) under accession number SRP000383.

References

1. Hausen Zur H: **Human papillomavirus & cervical cancer.** *Indian J Med Res* 2009, **130**:209.

2. Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, Moore PS: **Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma.** *Science* 1994, **266**:1865–1869.

3. Cover TL, Blaser MJ: **Helicobacter pylori in health and disease.** *Gastroenterology* 2009, **136**:1863–1873.

4. Polk DB, Peek RM: **Helicobacter pylori: gastric cancer and beyond.** *Nat. Rev. Cancer* 2010, **10**:403–414.

5. Goodman AL, Gordon JI: **Our Unindicted Coconspirators: Human Metabolism from a Microbial Perspective.** *Cell Metab* 2010, **12**:111–116.

6. Hope ME, Hold GL, Kain R, El-Omar EM: **Sporadic colorectal cancer--role of the commensal microbiota.** *FEMS Microbiol. Lett.* 2005, **244**:1–7.

7. Yang L, Pei Z: **Bacteria, inflammation, and colon cancer.** *World J. Gastroenterol.* 2006, **12**:6741–6746.

8. Rowland IR: **The role of the gastrointestinal microbiota in colorectal cancer.** *Curr. Pharm. Des.* 2009, **15**:1524–1527.

9. Sears CL, Pardoll DM: **Perspective: Alpha-Bugs, Their Microbial Partners, and the Link to Colon**

Cancer. *Journal of Infectious Diseases* 2011, **203**:306–311.

10. Bass AJ, Lawrence MS, Brace LE, Ramos AH, Drier Y, Cibulskis K, Sougnez C, Voet D, Saksena G, Sivachenko A, et al.: **Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion**. *Nat. Genet.* 2011, doi:10.1038/ng.936.

11. Kostic AD, Ojesina AI, Pedamallu CS, Jung J, Verhaak RGW, Getz G, Meyerson M: **PathSeq: software to identify or discover microbes by deep sequencing of human tissue.** *Nat. Biotechnol.* 2011, **29**:393–396.

12. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C: **Metagenomic biomarker discovery and explanation.** *Genome Biol.* 2011, **12**:R60.

13. Strauss J, Kaplan GG, Beck PL, Rioux K, Panaccione R, Devinney R, Lynch T, Allen-Vercoe E: **Invasive potential of gut mucosa-derived Fusobacterium nucleatum positively correlates with IBD status of the host.** *Inflamm. Bowel Dis.* 2011, **17**:1971–1978.

14. Loy A, Arnold R, Tischler P, Rattei T, Wagner M, Horn M: **probeCheck--a central resource for evaluating oligonucleotide probe coverage and specificity.** *Environ. Microbiol.* 2008, **10**:2894–2898.

15. Swidsinski A, Dorffel Y, Loening-Baucke V, Theissig F, Ruckert JC, Ismail M, Rau WA, Gaschler D, Weizenegger M, Kuhn S, et al.: **Acute appendicitis is characterised by local invasion with Fusobacterium nucleatum/necrophorum.** *Gut* 2011, **60**:34–40.

16. Neut C, Bulois P, Desreumaux P, Membre J-M, Lederman E, Gambiez L, Cortot A, Quandalle P, van Kruiningen H, Colombel J-F: **Changes in the bacterial flora of the neoterminal ileum after ileocolonic resection for Crohn's disease.** *Am J Gastroenterol* 2002, **97**:939–946.

17. Ohkusa T, Sato N, Ogihara T, Morita K, Ogawa M, Okayasu I: **Fusobacterium varium localized in the colonic mucosa of patients with ulcerative colitis stimulates species-specific antibody.** *J. Gastroenterol. Hepatol.* 2002, **17**:849–853.

18. Moore WE, Moore LV: **The bacteria of periodontal diseases.** *Periodontol 2000* 1994, **5**:66–77.

19. Bachrach G, Ianculovici C, Naor R, Weiss EI: **Fluorescence based measurements of Fusobacterium nucleatum coaggregation and of fusobacterial attachment to mammalian cells.** *FEMS Microbiol. Lett.* 2005, **248**:235–240.

20. Uitto V-J, Baillie D, Wu Q, Gendron R, Grenier D, Putnins EE, Kanervo A, Firth JD: **Fusobacterium nucleatum increases collagenase 3 production and migration of epithelial cells.** *Infection and Immunity* 2005, **73**:1171–1179.

21. Han YW, Shi W, Huang GTJ, Kinder Haake S, Park N-H, Kuramitsu H, Genco RJ: **Interactions between Periodontal Bacteria and Human Oral Epithelial Cells: Fusobacterium nucleatum Adheres to and Invades Epithelial Cells**. *Infection and Immunity* 2000, **68**:3140–3146.

22. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al.: **Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities.** *Appl. Environ. Microbiol.* 2009, **75**:7537–7541.

23. Ye Y: **Identification and Quantification of Abundant Species from Pyrosequences of 16S rRNA by Consensus Alignment.** *Proceedings (IEEE Int Conf Bioinformatics Biomed)* 2011, **2010**:153–157.

24. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO: **SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB.**

Nucleic Acids Res 2007, **35**:7188–7196.

25. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, et al.: **Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons.** *Genome Res* 2011, **21**:494–504.

26. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, et al.: **The Ribosomal Database Project: improved alignments and new tools for rRNA analysis.** *Nucleic Acids Res* 2009, **37**:D141–5.

27. Boutaga K, van Winkelhoff AJ, Vandenbroucke-Grauls CMJE, Savelkoul PHM: **Periodontal pathogens: A quantitative comparison of anaerobic culture and real-time PCR**. *FEMS Immunol. Med. Microbiol.* 2005, **45**:191–199.

CHAPTER 3

Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor immune microenvironment

This Chapter is a reproduction of a manuscript that has been submitted for publication. Detailed author information appears below:

Aleksandar D. Kostic[1-3], Eunyoung Chun[4], Lauren Robertson[4], Jonathan N. Glickman[5,6], Carey Ann Gallini[4], Monia Michaud[4], Thomas E. Clancy[1,2,5], Daniel C. Chung[1,7], Paul Lochhead[8], Georgina L. Hold[8], Emad M. El-Omar[8], Dean Brenner[9], Charles S. Fuchs[1,2,4], Matthew Meyerson[1-3*] and Wendy S. Garrett[1-4*]

[1]Departments of Medicine, Pathology, and Surgery, Harvard Medical School, Boston, MA 02115, USA
[2]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA
[3]Cancer Program, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
[4]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115, USA
[5]Department of Surgery, Brigham and Women's Hospital, Boston, MA 02115, USA
[6]Miraca Life Sciences, Inc. Newton, MA 02464, USA
[7]Department of Medicine, Massachusetts General Hospital, Boston MA, 02114
[8]School of Medicine and Dentistry, University of Aberdeen, Aberdeen, Scotland AB25 2ZD, United Kingdom
[9]Cancer and Geriatrics Center, University of Michigan Medical Center, Ann Arbor, MI 48109, USA
[*]indicates equal contribution

Author Contributions:

A.D.K. and W.S.G. designed and carried out experiments, analyzed data and wrote the manuscript. M.M. designed the experiments, analyzed data and wrote the manuscript. E.C. carried out experiments and assisted in writing the manuscript. L.R., C.G., M. Michaud performed experiments. J.N.G. performed all histological assessment and was blinded to the experimental intervention. T.E.C., D.C., P.L., G.L.H., E.M.E., D.B. and C.S.F. provided human samples.

Abstract

Deep-sequencing metagenomic analyses have revealed that *Fusobacterium* spp. are associated with human colorectal carcinoma. Here we show that in the $Apc^{Min/+}$ mouse model of intestinal tumorigenesis, *Fusobacterium nucleatum* increases tumor multiplicity, selectively recruits tumor-infiltrating myeloid cells, and is associated with a pro-inflammatory expression signature that is shared with human fusobacteria-positive colorectal carcinomas. However, in contrast with recent studies of other bacteria in colorectal carcinoma, *F. nucleatum* does not exacerbate colitis or colitis-associated colorectal carcinoma. We find that *Fusobacterium* spp. are enriched in human colonic adenomas relative to surrounding tissues and fusobacterial abundance is increased in stool samples from patients with colorectal adenomas and carcinomas, compared to healthy subjects. Collectively, these data support that fusobacteria may be involved in early stages of intestinal tumorigenesis and, through recruitment of tumor-infiltrating immune cells, may generate a pro-inflammatory tissue microenvironment conducive to colorectal neoplasia progression.

Introduction

There is accumulating evidence that members of the gut microbiota contribute to colorectal cancer, the second most incident cancer worldwide with over 1.2 million new cases per year [1]. The majority of studies have focused on a small subset of colorectal cancers, colitis-associated colorectal cancers, and employed rodent pre-clinical models. Antibiotic-treatment or absence of the gut microbiota reduced tumor incidence in several mouse colitis-associated colorectal cancer models [2-4]. Recently, two bacterial pathogens have been identified that

promote colitis-associated colorectal cancer. Enterotoxigenic *Bacteroides fragilis* induces colitis and colonic tumors in $Apc^{Min/+}$ mice by triggering a T helper type 17 (Th17) inflammatory response [5], while adherent-invasive *Escherichia coli* strain NC101 promotes colitis-associated colorectal cancer in monocolonized, azoxymethane-injected $Il10^{-/-}$ mice [6]. However, the majority of human colorectal cancers do not arise in the setting of inflammatory bowel disease.

The gut microbiota may be a driver in colorectal cancers that are not associated with colitis. Metagenomic analyses using whole-genome sequencing [7], transcriptome sequencing [8], or bacterial 16S ribosomal RNA gene DNA sequencing [7,9] have shown enrichment of *Fusobacterium* species in colorectal cancers relative to adjacent normal tissue. However, to date it has been unclear whether these findings represent an indirect association, or whether *Fusobacterium* spp. functionally contributes to colorectal cancer (CRC) tumorigenesis [10].

Here, we analyze the impact of *Fusobacterium nucleatum* on colorectal cancer progression and on tumor inflammation in mouse models, and assess the association of *F. nucleatum* in stool and colonic tissue from patients with colorectal adenomas and adenocarcinomas. Our data are consistent with the possibility that *F. nucleatum* potentiates non-colitis associated colorectal tumorigenesis.

Results

Fusobacterium nucleatum promotes intestinal tumorigenesis

The enrichment of *Fusobacterium* spp. in colorectal cancers observed in three independent human cohort studies [7-9] prompted us to examine whether *Fusobacterium* could accelerate tumorigenesis in mouse models. We began with pilot experiments in mice which have

a propensity to develop intestinal tumors because of a T to A transversion of nucleotide 2549 in one copy of the tumor suppressor gene *Apc* (C57BL/6 $Apc^{Min/+}$) or because of genetic defects resulting in chronic intestinal inflammation (BALB/c $Il10^{-/-}$ and BALB/c $T\text{-}bet^{-/-}$ X $Rag2^{-/-}$). We introduced a human clinical isolate of *Fusobacterium nucleatum*, several human clinical isolates of *Streptococcus* species (*S. anginosus* [4 mice]; *S. parasanguinis* [4 mice]; and *S. sanguinis* [4 mice]), or tryptic soy broth to $Apc^{Min/+}$ mice and *F. nucleatum* or tryptic soy broth to $IL10^{-/-}$ and $T\text{-}bet^{-/-}$ X $RAG2^{-/-}$ mice starting at 6 weeks of age. Streptococci were used as control strains because of the longstanding association of streptococcal species with occult colonic malignancies [11].

Introduction of *F. nucleatum* into $Apc^{Min/+}$ mice was associated with an accelerated onset of colonic tumors. $Apc^{Min/+}$ mice fed *F. nucleatum* developed a significantly higher number of colonic tumors at 3.5 months of age as compared to $Apc^{Min/+}$ mice fed *Streptococcus* (*Strep.*) spp., or tryptic soy broth control ($P < 0.001$, $P < 0.0001$) (**Fig. 3-1a-c**). However, *F. nucleatum* did not induce colitis (**Fig. 3-1a-c**) in contrast with enterotoxigenic *Bacteroides fragilis*, which causes colitis and accelerates tumorigenesis in $Apc^{Min/+}$ mice [5]. Also, *F. nucleatum* neither exacerbated intestinal inflammation nor accelerated tumorigenesis in the two mouse models of colitis-associated colorectal cancer, $Il10^{-/-}$ and $T\text{-}bet^{-/-}$ X $Rag2^{-/-}$, examined (**Fig. 3-1a**).

In addition to increased colonic tumor numbers, $Apc^{Min/+}$ mice fed *F. nucleatum* had a higher count of small intestinal adenomas versus *Strep.* spp ($P = 0.0002$), or soy broth ($P = 0.02$) and small intestinal adenocarcinomas versus *Strep.* spp ($P = .0017$) or tryptic soy broth ($P = .0082$) (**Fig. 3-1d, e**). Consistent with prior findings in human colorectal carcinoma [7,8], *F. nucleatum* was culturable from $Apc^{Min/+}$ tumors and enriched in tumor tissue relative to adjacent normal tissue in *F. nucleatum* fed $Apc^{Min/+}$ mice as assayed by qPCR (**Fig. 3-1f**). Furthermore, *F.*

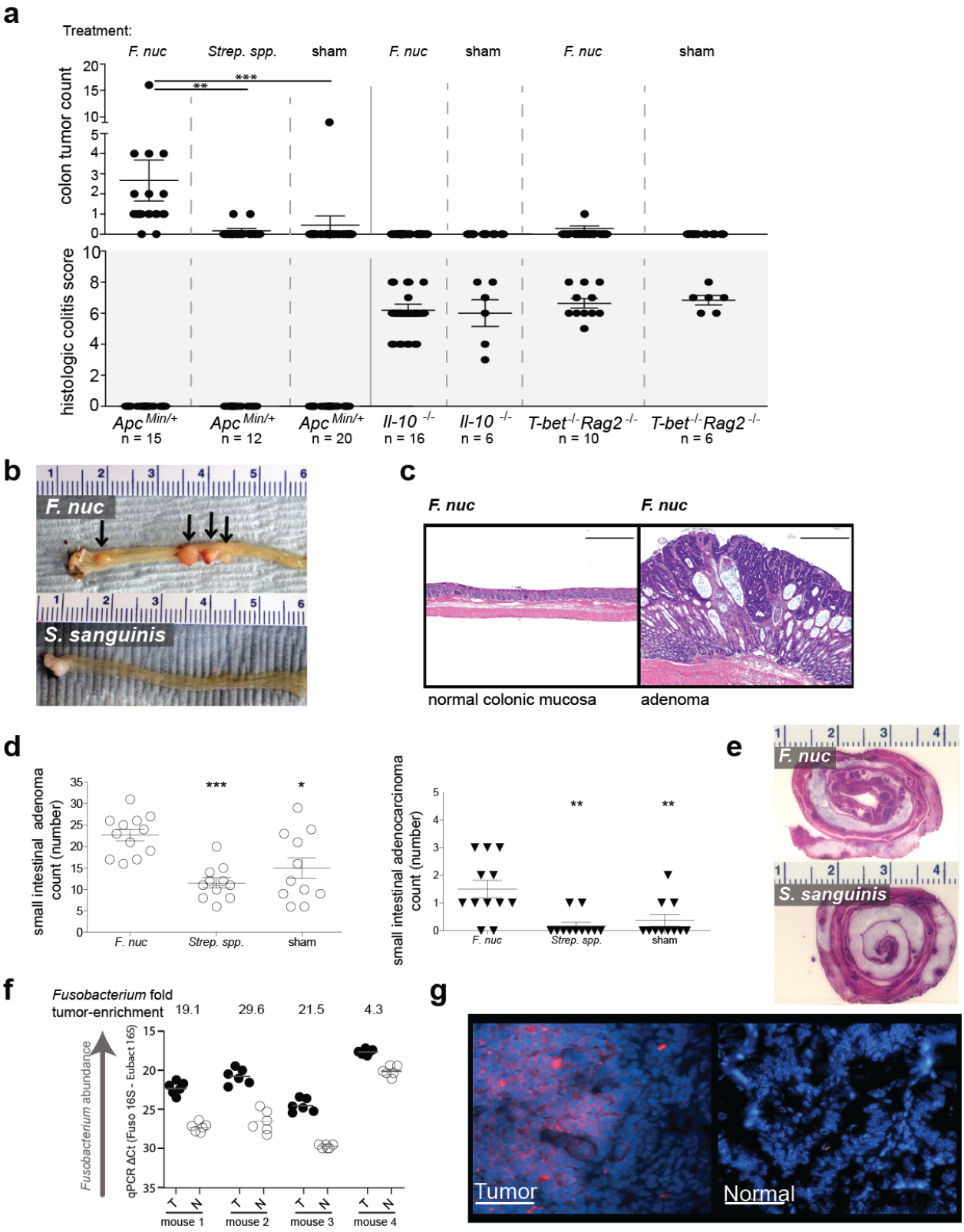


Figure 3-1. ***Fusobacterium nucleatum*** **promotes intestinal tumorigenesis and is enriched in tumor tissues of** ***Apc*** $^{Min/+}$ **mice.** (**a**) Gross colon tumor counts and histologic colitis scores from *Apc* $^{Min/+}$, *Il10* $^{-/-}$, and *T-bet* $^{-/-}$ X *Rag2* $^{-/-}$ mice fed *F. nucleatum* (*F. nuc*), *Streptococcus* spp., or soy broth control. Mice were started on the 8-week, daily

(*Figure 3-1 Continued)* feeding regimen at 6 weeks of age. (*** $P < 0.0001$, ** $P < 0.001$, * $P < 0.01$). (**b**) Representative images of gross colons (ruler numbers in cm) and (**c**) colonic histological analysis of $Apc^{Min/+}$ mice (100 micrometer scale bar). (**d**) Histopathologic small intestinal adenoma and adenocarcinoma counts in $Apc^{Min/+}$ mice. (**e**) Representative sections of rolled small intestines from $Apc^{Min/+}$ mice (ruler numbers in cm). (**f**) *Fusobacterium* abundance in matched tumor (T) versus normal (N) tissues from colons of $Apc^{Min/+}$ mice fed *F. nucleatum* measured by quantitative PCR. (**g**) Representative FISH images of tumor and matched normal tissue from an $Apc^{Min/+}$ mouse fed *F. nucleatum* using a *Fusobacterium* 16S rDNA-directed probe (50 micrometer scale bar).

---

nucleatum had a broad distribution within tumor tissue when visualized with fluorescence *in situ* hybridization using a *Fusobacterium* 16S rRNA-directed probe (**Fig. 3-1g**). Together, these results indicate that *F. nucleatum* may accelerate tumorigenesis in the absence of colitis, or macroscopic inflammation, in $Apc^{Min/+}$ mice.

F. nucleatum selectively expands myeloid-derived immune cells

Immune cells and their effectors have been found to be key components of the tumor milieu that promote neoplastic progression [12]. This type of inflammation has been referred to as intrinsic inflammation, as it is intrinsic to the tumor, in contrast with the extrinsic inflammation of IBD and colitis-associated colorectal cancer [12]. To address whether *F. nucleatum* contributes to tumorigenesis by an intrinsic inflammatory mechanism, we characterized and quantified the tumor infiltrating immune cells from the intestinal tumors of $Apc^{Min/+}$ mice that were fed *F. nucleatum* or *Streptococcus sanguinis* for 8 weeks, or were not fed bacteria over the same time frame. Small intestinal rather than colonic tumors were used, because controls did not develop sufficient numbers of colon tumors within the experimental timeframe.

We observed a striking increase in infiltrating cells of the myeloid lineage in the tumors from *Fusobacterium*-treated mice. $CD11b^+$ myeloid cells (mean 3.4X higher in cell number, mean 4.0X higher in % population) and $CD11c^+$MHC class II^+ dendritic cells (DC) (mean 3.1X higher in cell number, mean 2.7X higher in % population) increased in the tumors of $Apc^{Min/+}$

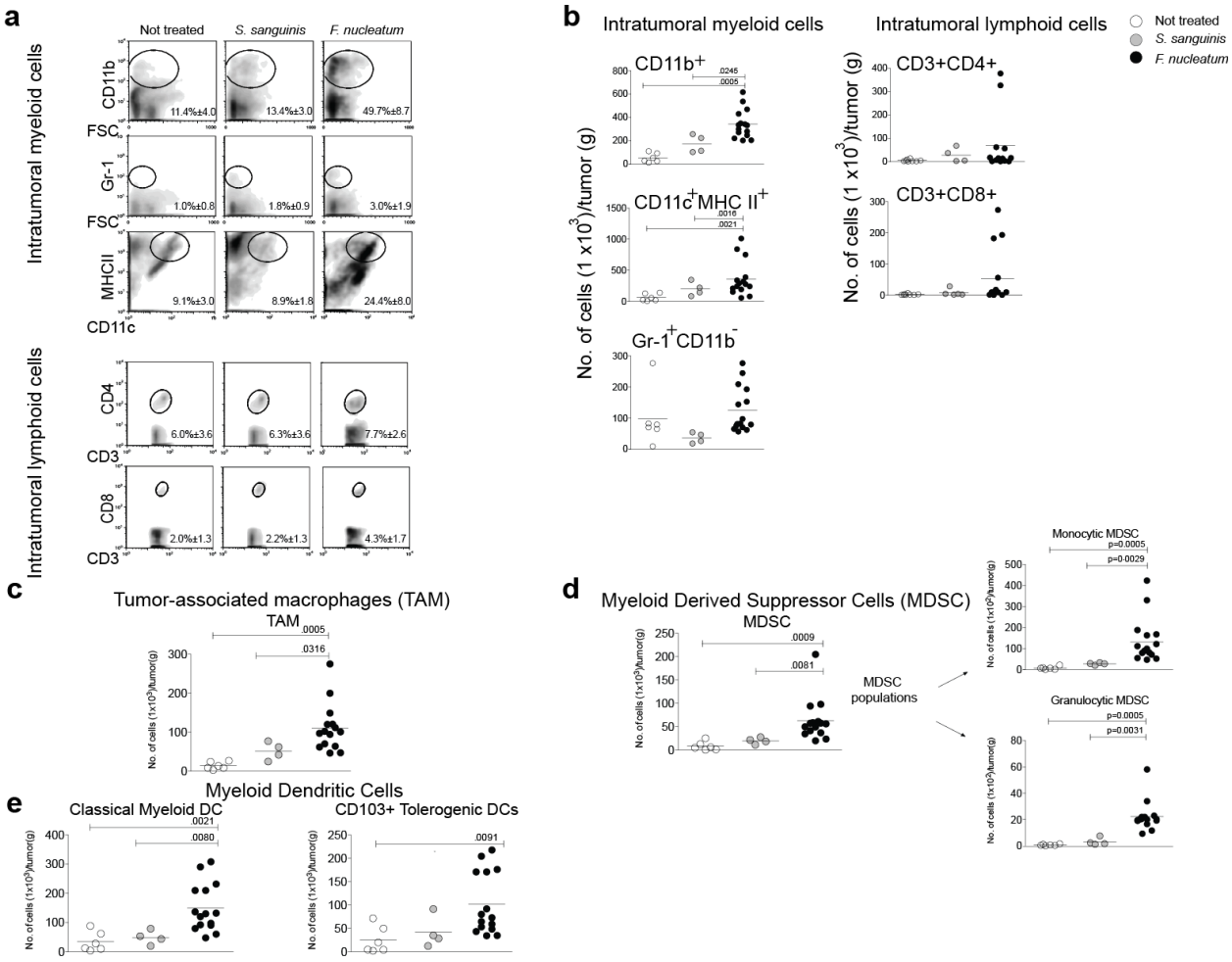


Figure 3-2. *F. nucleatum* selectively expands myeloid-derived immune cells, but not lymphoid immune cells in the intestinal tumor microenvironment. Flow cytometric analyses: (**a**) Percentage of intratumoral myeloid cells (upper panel) and lymphoid cells (lower panel). $CD11b^+$ myeloid cells, $CD11c^+MHCII^{hi}$ dendritic cells, $Gr\text{-}1^+$ $CD11b^-$ granulocytic neutrophils, $CD3^+CD4^+$ T cells, or $CD3^+CD8^+$ T cells (y-axes) vs forward scatter (FSC). Mean percentages ± s.e.m are shown within each plot. n= 6, 4, or 15 for not treated, *S. sanguinis,* or *F. nucleatum*, respectively. (**b**) Cell number/gram tumor for myeloid and lymphoid cells from the treatment groups. (**c**) Cell number/gram tumor for TAM ($CD45^+CD11b^+F4/80^+$). (**d**) Left panel, cell number/gram tumor for MDSC ($CD45^+CD11b^+Gr\text{-}1^+$). Right upper panel, cell number/gram tumor of monocytic MDSC ($CD45^+CD11b^+Gr\text{-}1^{int}$ $Ly6C^{hi}$) and lower panel, granulocytic MDSC ($CD45^+CD11b^+Gr\text{-}1^{hi}Ly6C^{low}$). (**e**) Left panel, cell number/gram tumor of classical myeloid DCs ($CD45^+CD11c^+MHCII^{hi}$ $CD11b^+$) and right panel, cell number/gram tumor of tolerogenic DCs ($CD45^+CD11c^+MHCII^{hi}CD103^+$). Each symbol represents data from an individual mouse. *P* values are shown where significant.

mice fed invasive *F. nucleatum* as compared to controls (**Fig. 3-2a, b**). In contrast, numbers of intratumoral $CD11b^{-}Gr\text{-}1^{+}$ granulocytic neutrophils and T lymphocytes, $CD3^{+}CD4^{+}$ and $CD3^{+}CD8^{+}$ cells, in mice fed *F. nucleatum* were not significantly different from controls (**Fig. 3-2a, b**).

Substantial experimental data from clinical and pre-clinical studies indicate that tumor-associated macrophages (TAM) promote tumor progression and metastasis [13-15]. We further characterized the intratumoral myeloid populations and found an enrichment of TAMs ($CD11b^{+}F4/80^{+}$) in *F. nucleatum* fed mice as compared to controls (mean 4.1X increased cell number) (**Fig. 3-2c**). Myeloid-derived suppressor cells (MDSCs) represent an additional population of tumor-permissive myeloid cells with potent immune suppressive activity [16-19]. MDSCs ($CD11b^{+}Gr\text{-}1^{+}$) were enriched (mean 4.9X increased cell number) in *F. nucleatum* fed mice vs controls. There are two principal MDSC subsets, monocytic and granulocytic, both of which increased in the tumors of *F. nucleatum* fed mice (monocytic MDSCs ($CD11b^{+}Gr\text{-}1^{int}Ly6C^{hi}$), mean 9.1X and granulocytic MDSCs ($CD11b^{+}Gr\text{-}1^{hi}Ly6C^{low}$), mean 11.6X increased cell number) compared to controls (**Fig. 3-2d**).

Within tumors, dendritic cells (DC) can either dampen or promote anti-tumor immunity [20]. The goal of several current tumor immunotherapy efforts is to re-program and activate DCs within tumors, as intratumoral DCs can assume tumor permissive phenotypes [21]. The intestine possesses a specific subset of tolerance promoting DCs. These tolerogenic DCs express the cell surface integrin CD103 and promote the expansion of $Foxp3^{+}$ regulatory T cells, a $CD4^{+}$ T cell subset that suppresses cytotoxic and effector T cells and thus dampen anti-tumor immunity [22-24]. Having identified an expansion of DCs in *F. nucleatum* fed mice, we further characterized the DC populations and found increases in classical myeloid DCs ($CD11c^{+}MHCII^{+}CD11b^{+}$) and

$CD103^+$ tolerogenic DCs ($CD11c^+MHCII^{hi}CD103^+$) (mean 3.5X and mean 2.8X increased cell number, respectively) compared to controls. (**Fig. 3-2e**). To determine whether the changes in TAMs, MDSCs, and DCs were impacting specific subsets of $CD4^+$ T cells implicated in colon tumorigenesis [5,25], we examined both intratumoral T-helper 17 (Th17) cells and $CD4^+$ $Foxp3^+$ regulatory T cells. While there was a trend towards an increase in $Foxp3^+$ regulatory T cells (Treg) and Th17 cells in *F. nucleatum* fed mice, there was significant heterogeneity within this group, that did not correlate with intratumoral *F. nucleatum* abundance, and differences were not statistically significant (**Supplementary Fig. 3-1).** Additionally, there was no trend between Th17 and Treg cell numbers in the same individual mice. Collectively, these data support that *F. nucleatum* modulates the tumor immune microenvironment and results in expansion of selective myeloid-derived immune cell types that have been well-described to promote tumor progression [14,26,27].

A *Fusobacterium*-associated human colorectal cancer gene signature shared and validated in mice

Given our findings of *F. nucleatum*-induced myeloid-derived cell expansion in mouse intestinal tumors, we asked if there would also be a similar immunological profile in human *Fusobacterium*-associated colon tumor transcriptomic data. We utilized a data set of deep transcriptome sequencing (i.e. RNA-Seq) of 133 colon tumors generated by The Cancer Genome Atlas [28]. All non-human sequencing reads were applied to PathSeq, a computational tool that identifies and quantifies the abundance of all bacteria present in each tumor [29]. By calculating the Spearman's rank correlation coefficient of the relative abundance of *Fusobacterium* spp. transcripts with host gene expression in this dataset, we identified a *Fusobacterium*-associated human CRC gene expression signature. We found a correlation of immune cell marker genes

associated with TAM (*CD209*, *CD206/MRC1*, *IL-6*, *IL-8*, and *CXCL10*), MDSC (*CD33* and *IL-6*), and DC (*CD11c/ITGAX*, *CD209*, *TNF*, *CD80*) with *Fusobacterium* abundance in human *Fusobacterium*-associated tumors by RNA-Seq, similar to our findings with *Fusobacterium* in mouse flow cytometry experiments (**Fig. 3-3a**). Analysis using the Ingenuity IPA gene ontology module revealed that the *Fusobacterium*-associated human CRC gene expression signature was highly enriched for the inflammatory response gene ontology category (corrected $P < 1x10^{-33}$; Ingenuity analysis) as well as other categories also related to immune and inflammatory disease (**Supplementary Table 3-1**). To test the specificity of this correlation, we performed the same analysis with the other top-four highly abundant genera besides *Fusobacterium* (*Bacteroides, Escherichia, Streptococcus, and Propionobacterium*), but none of these other genera had enrichment for inflammation-related gene functions, nor did they have comparably high gene ontology enrichments in any other functional categories (**Supplementary Fig. 3-2**). Relative abundance of *Fusobacterium spp*. transcripts are shown for each of the 133 TCGA colon tumors with scaled expression values for the top 50 ranked genes denoted as the row Z-score in a heat map (**Fig. 3-3b**). Many of the these top ranked genes, *PTGS2 (COX-2), IL1β*, *IL6*, *IL8*, and *TNF* (*TNF-α)*, have not only been investigated in colorectal carcinogenesis but also are induced by *Fusobacterium* in co-culture with human and mouse cell lines *in vitro* [30-32].

The expression signature was suggestive of a NF-κB-driven pro-inflammatory response [33]. As NF-κB has been identified as a central link between inflammation and cancer [34], we assessed whether there was a correlation between increased NF-κB activation and *Fusobacterium* abundance in human colorectal cancers. We obtained freshly resected human colorectal cancer samples and generated nuclear extracts from these samples. After stratifying the samples by their abundance of *Fusobacterium* spp., we performed western blots on the

a

Correlation with *Fusobacterium* abundance	Spear-man	corrected *P*-val
Macrophages		
CD209	0.290	0.0343
MRC1 (CD206)	0.269	0.0446
IL6	0.419	0.0007
IL8	0.423	0.0007
CXCL10	0.261	0.0499
MDSCs		
CD33	0.302	0.0286
IL6	0.419	0.0007
Dendritic cells		
ITGAX (CD11c)	0.294	0.0329
CD209	0.290	0.0343
TNF	0.346	0.0097
CD80	0.326	0.0174

b

Fusobacterium-associated host gene expression signature	corrected *P*-val
GO Biological Function	
Inflammatory response	2.9×10^{-34}
Cell-cell signaling & interact.	2.0×10^{-27}
Immune cell trafficking	2.4×10^{-24}
Cellular growth & proliferation	1.1×10^{-24}
Cell death	1.3×10^{-21}
Immunological disease	3.1×10^{-17}
Inflammatory disease	3.1×10^{-17}
Lymphoid tissue struct. & dev.	5.5×10^{-16}

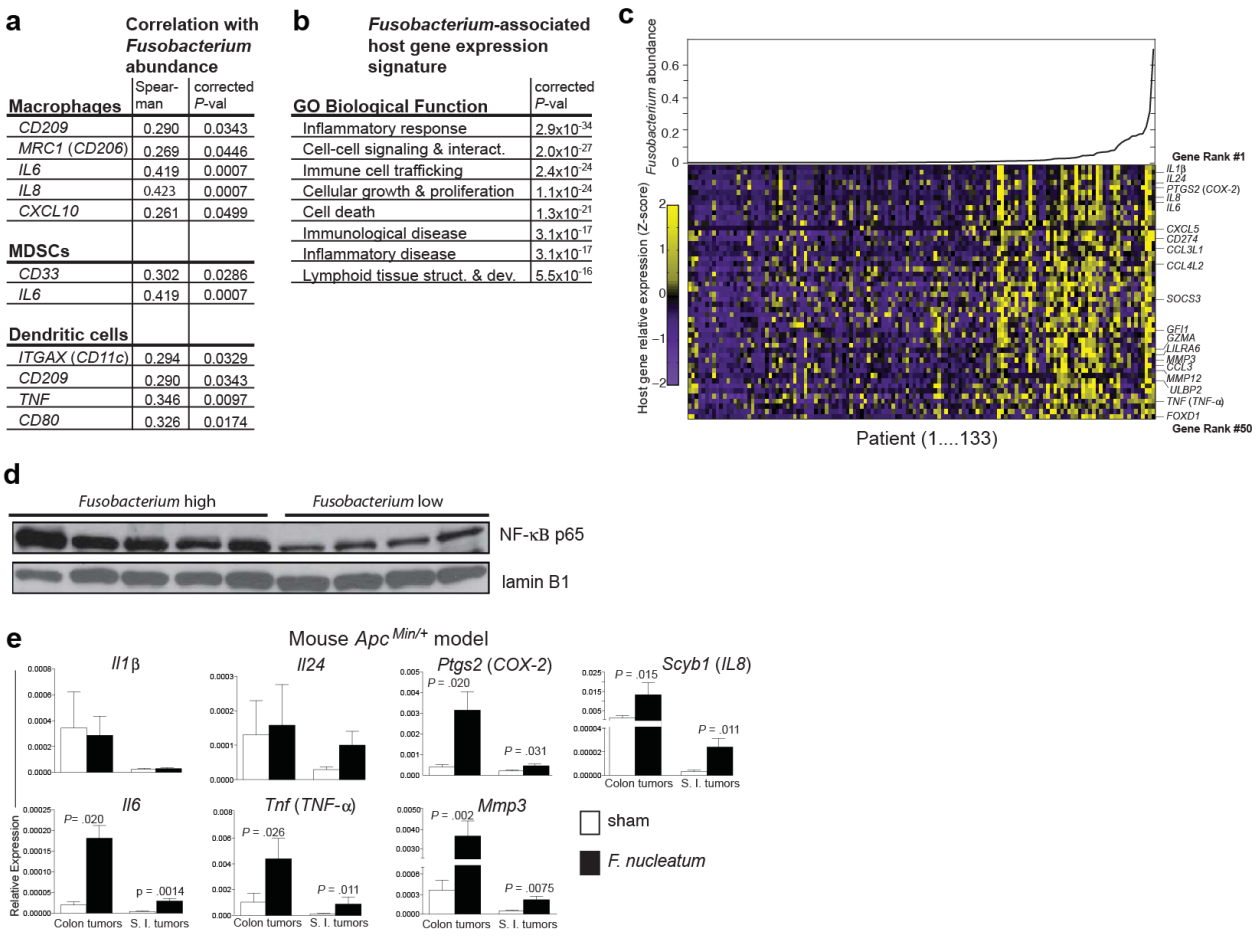


Figure 3-3. A *Fusobacterium*-associated human colorectal cancer gene signature shared and validated in mice. (**a**) Immune cell types enriched in *Fusobacterium*-associated mouse tumor are shown with the human marker gene utilized to determine their abundance in the TCGA CRC RNA-seq data set. Spearman's rank correlation coefficient of the relative abundance of *Fusobacterium* spp. transcripts and *P* values are shown next to each marker gene. (**b**) Ingenuity Pathway Analysis Biological Function Gene Ontology categories that are enriched for *Fusobacterium* abundance-correlating gene sets are shown. (**c**) Relative abundance of *Fusobacterium* spp. transcripts are plotted for each of the 133 TCGA colon tumors (upper panel and lower panel x-axis) and scaled expression values for the top 50 ranked genes denoted as the row Z-score (y-axis) are shown in a heat map (lower panel) with a purple (low expression)–yellow (high expression) color scale. (**d**) Western blot of nuclear extracts from human colon cancer with a high or low *Fusobacterium* relative abundance (see Methods) using an anti-NF-kB p65 antibody and the nuclear marker lamin B1. (**e**) qPCR analysis of a selection of the top 50 ranked genes in (B) in colon and small intestinal tumors from *F. nucleatum* vs tryptic soy broth fed $Apc^{Min/+}$ mice. Tumors from 6-9 mice per group were used. Means and s.e.m. are plotted. *P* values are shown where significant.

nuclear extracts from *Fusobacterium* high and low tumors to examine NF-κB activation. NF-κB was indeed more activated (increased nuclear translocation of the p65 NF-κB subunit) in tumors with a high vs low *Fusobacterium* abundance (**Fig. 3-3c, upper panel**).

We found that most of the human *Fusobacterium*-associated pro-inflammatory genes with mouse homologs including: *Ptgs2* (*COX-2* mouse homolog), *Scyb1* (*IL8* mouse homolog), *Il6*, *Tnf* (*TNFa*), and *Mmp3* were also more highly expressed in both small intestinal and colonic tumors from mice that were treated with *F. nucleatum* vs tryptic soy broth (**Fig. 3-3e**).

Fusobacteria are enriched in colonic adenomas and in stools samples from patients with adenomas and colorectal carcinomas

Colonic adenomas are neoplastic epithelial lesions that have the potential to become malignant and are believed to be the precursors of the majority of sporadic colorectal cancers. Given the enrichment of *Fusobacterium* spp. in colorectal tumor versus adjacent normal tissue, we examined if there would be a similar enrichment in adenomas, suggesting the involvement of *Fusobacterium* in neoplastic initiation or progression, prior to the establishment of carcinoma. A few recent studies have performed case control studies of colorectal adenomas using 16S rRNA gene surveys with one noting fusobacterial enrichments in their patient cohort [35-38]. We measured *Fusobacterium* spp. abundance in paired adenoma tissue versus adjacent normal tissue from the same patient drawing samples from several geographic locations and registries: The Cooperative Human Tissue Network (Eastern, Southern, and Western divisions of the United States), Massachusetts General Hospital (Boston, MA), and University of Aberdeen School of Medicine (Aberdeen, United Kingdom). We found that *Fusobacterium* was detectable by qPCR in 48% of adenomas (n=29), and in those cases that were positive, *Fusobacterium* was enriched in adenomas relative to surrounding tissue ($P < 0.004$) (**Fig. 3-4a** and **Supplementary Table 3-**

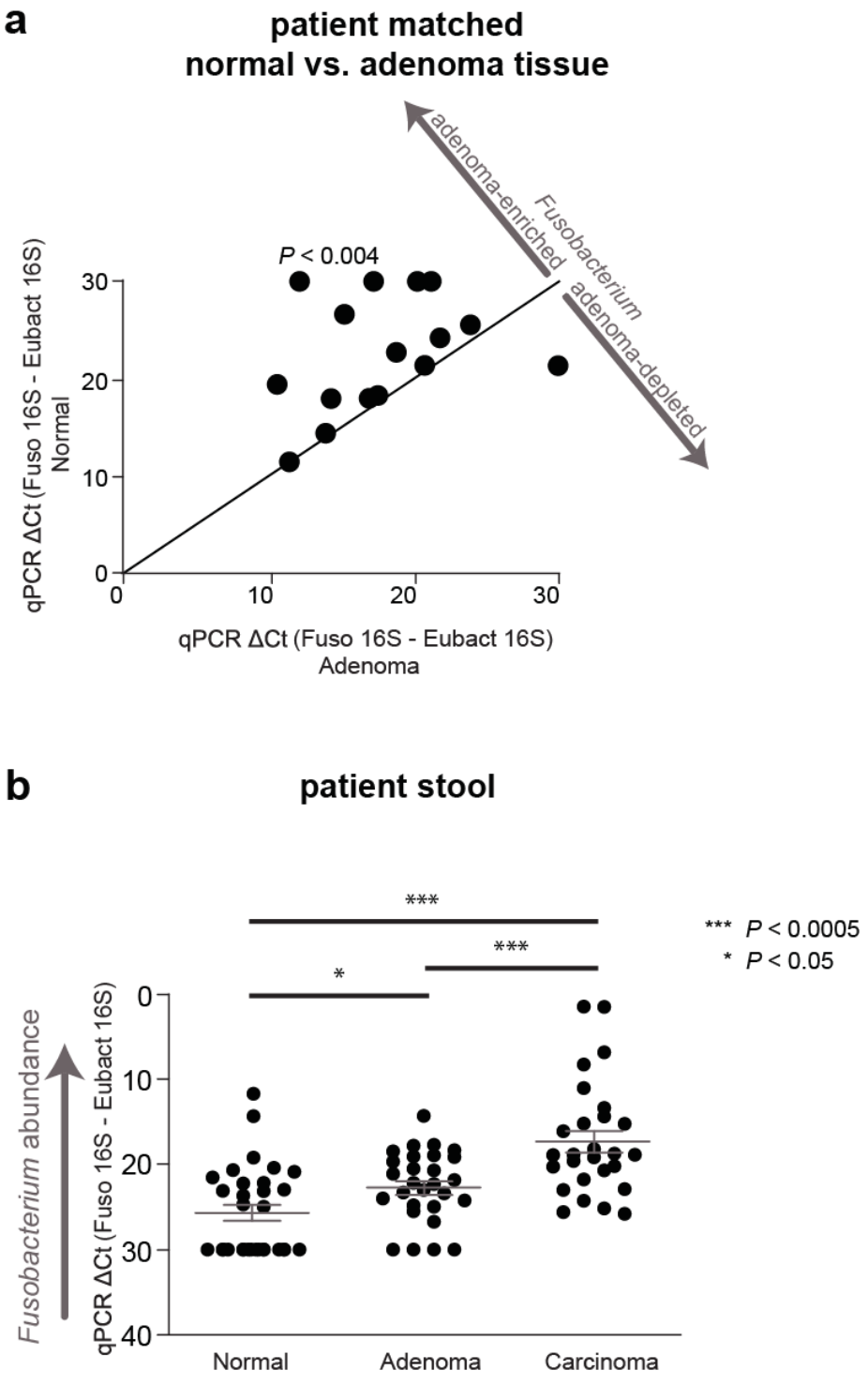


Figure 3-4. *Fusobacterium* is enriched in adenoma versus adjacent normal tissue and detected at a higher abundance in stool from CRC and adenoma cases than from healthy controls. (**a)** *Fusobacterium* abundance for normal tissue (x-axis) vs adenoma (y-axis) is plotted. Samples with no difference in fusobacterial abundance between adenoma and normal within a single patient are plotted on the diagonal line. 29 matched adenoma normal tissues pairs were tested. Each symbol represents data from one patient (adenoma and normal tissue). (**b)** Fecal *Fusobacterium* abundance from healthy subjects (n=30), subjects with colorectal adenomas (n=29), and colorectal cancer (n = 27).

3). These data suggest that *Fusobacterium* begins to accumulate at early stages of colonic tumorigenesis in some groups of patients.

Next, we determined if *Fusobacterium* spp. are enriched exclusively in tumor and adenoma tissue, or whether fusobacteria have a higher overall abundance in the fecal microbiota of CRC patients relative to healthy controls in a case-control experiment. We compared the levels of *Fusobacterium* spp. to universal Eubacteria 16S by quantitative qPCR of stool. These subjects provided their stool samples to the Early Detection Research Network (EDRN) prior to bowel preparation and screening colonoscopy and had no prior history of colorectal cancer or gastrointestinal disease (**Fig. 3-4b** and **Supplementary Table 3-4**). We found that *Fusobacterium* spp. were enriched in CRC patients ($P < 1x10^{-5}$). Because of the *Fusobacterium* enrichment in adenomas (**Fig. 3-4a**), we examined the *Fusobacterium* spp. abundance in stool from patients who provided stools to the EDRN and who were subsequently diagnosed with colonoscopy-confirmed colonic adenomas (**Fig. 3-4b** and **Supplementary Table 3-4**). We found that *Fusobacterium* spp. were also enriched in subjects with adenomas as compared to healthy control individuals ($P < 5x10^{-3}$). These results indicate that increased abundance of *Fusobacterium* species in the gut microbiota may be a general feature of colonic tumorigenesis.

Discussion

Entry of microbes and microbial products into the evolving tumor microenvironment potentiates tumor growth by eliciting tumor-promoting immune cell responses [25,39]. Our results demonstrate that *Fusobacterium* spp., rare gut microbiome constituents in the healthy human population [40], are found at increased abundance in the stool of patients with adenomas

and colorectal cancer, and are enriched in adenomas and adenocarcinomas relative to non-involved colonic tissues. Introduction of *Fusobacterium nucleatum* to $Apc^{Min/+}$ mice resulted in accelerated small intestinal and colonic tumorigenesis, infiltration of specific myeloid cell subsets into tumors, and an NF-κB pro-inflammatory signature. This pro-inflammatory signature was shared with human colorectal cancer tissue with a high *Fusobacterium* abundance.

In contrast to other bacterial-driven models of intestinal tumorigenesis [5,6], we have found a specific bacterial strain that accelerates intestinal tumorigenesis in the absence of colitis. While $Apc^{Min/+}$ mice fed *F. nucleatum* exhibited enhanced intestinal tumorigenesis, neither $Il10^{-/-}$ nor *T-bet*$^{-/-}$ X *Rag2*$^{-/-}$ mouse models of colitis showed accelerated tumorigenesis or exacerbated colitis upon introduction of *F. nucleatum*. This may suggest that the tumorigenic effects of fusobacteria operate downstream of the loss of the tumor suppressor *APC* and the resulting intestinal dysplasia that occurs in $Apc^{Min/+}$ mice. This is relevant to most cases of human CRC, as only 2% of CRC cases are linked to colitis, but greater than 80% of non-hypermutated CRC tumors bear *APC* mutations [28]. Our *Fusobacterium* findings are also relevant to adenomas because mutations in *APC* are among the earliest molecular alterations that occur in an epithelium as it transitions to become an adenoma [41]. Therefore, early tumor-initiating somatic mutations likely precede the tissue enrichment of *Fusobacterium* spp. These mutations may contribute to the development of epithelial barrier defects, featuring the loss of tight junctions, cell-to-cell contacts, epithelial polarity and the mucus layer [25]. Intestinal barrier defects at local sites of dysplasia may promote the infiltration of *Fusobacterium* spp., among other bacteria and microbial products, allowing fusobacteria to take up residence in the tumor environment. This may represent a crucial stage in colorectal neoplasia wherein myeloid cell-mediated

immune responses provide the driving force for inflammatory genotoxic and epigenetic changes that lead to cancer.

Although barrier defects expose the intestinal mucosa to the entire luminal microbial milieu, *Fusobacterium* spp. become the most highly enriched bacterium in colorectal tumors relative to adjacent tissue [7,8,10]. This enrichment may be attributable to the strong adhesive and invasive abilities of fusobacteria to epithelial cells [42,43]. The tumor enrichment of fusobacteria may also result from the growth advantage it provides to the tumor by eliciting pro-tumorigenic responses from myeloid immune cells. Alternatively, fusobacterial metabolic specializations may endow it with a competitive advantage in the evolving tumor milieu. *Fusobacterium nucleatum* is an asaccharolytic bacterium so, unlike the *Enterobacteriaceae,* it will not compete for glucose, a preferred substrate for tumor metabolism [44]. Instead fusobacteria can utilize amino acids and peptides as nutrient sources in the tumor microenvironment. Products of amino acid metabolism generated by fusobacteria*,* including formyl-methionyl-leucyl-phenylalanine and short chain fatty acids, are myeloid cell chemoattractants, which may explain the intratumoral myeloid cell expansion we observed and interconnect tumor metabolism, bacterial metabolism, and immune cell function within the tumor microenvironment.

In addition, *F. nucleatum* strains, unlike many strict anaerobes of the intestinal lumen, possess a rudimentary electron transport chain, endowing them with a limited ability to respire oxygen [45]. Thus, *F. nucleatum* may be able to persist and slowly replicate in the hypoxic tumor microenvironment. Adhesive molecules that contribute to invasivity in *F. nucleatum* can promote bacterial aggregation and biofilm formation that also enhance oxygen tolerance [46]. Products of fusobacterial metabolism may make the tumor microenvironment more tumor-

permissive over time by directly promoting tumor cell proliferation, blood vessel growth, or immune cell infiltration.

We have shown that, in both human and mouse intestinal tumors, the pro-inflammatory gene expression signature associated with *Fusobacterium* features the up-regulation of *PTGS2* (*COX-2*). Epidemiological and clinical data suggest that non-steroidal anti-inflammatory drugs (NSAIDs) may be effective as a primary and secondary preventative measure in colorectal neoplasia [47]. Our findings on *Fusobacterium* spp. and the intrinsic inflammation it elicits may further explain why anti-inflammatory strategies such as NSAIDs are an effective colorectal cancer prevention strategy.

If our results demonstrating that fusobacteria potentiate tumorigenesis can be extended to human CRC, then targeting *Fusobacterium* populations in the oral cavity where it is most abundant [48], or in the gastrointestinal tract may work to delay or prevent tumor progression in patients at increased risk for CRC. We show that *Fusobacterium* is absent, or present at low basal levels, in the stool of healthy individuals, but is significantly more abundant in patients with a colonoscopy-confirmed adenoma or CRC. These findings suggest a need for epidemiological studies to evaluate whether *Fusobacterium* abundance can be utilized as a prognostic or diagnostic factor in colorectal cancer.

Materials and Methods

Bacterial strains and culturing

Fusobacterium nucleatum (EAVG_002; 7/1) [43] was a gift from the laboratory of Emma Allen-Vercoe (U. Guelph). *Streptococcus* species were isolated from freshly resected human

colon tissues and identified by a combination of Gram stain and morphology, catalase test, hemolysis on Difco™ Tryptic Soy Blood Agar Base No. 2 (BD, Sparks, MD) supplemented with 5% defibrinated sheep blood (Northeast Laboratory, Waterville, ME), and Sanger sequencing on 16S rDNA positions 27-1492.

Bacteria were stored at -80°C in autoclaved freezing media, consisting of 1% DMSO, 1% glycerol, and 12% w/v skim milk powder in distilled water. *F. nucleatum* strains were plated on Fastidious Anaerobe Agar (Neogen, Lansing, MI) supplemented with 5% defibrinated sheep blood (Northeast Laboratory, Waterville, ME) and grown overnight in Bacto™ Tryptic Soy Broth (BD, Sparks, MD) (TSB) supplemented with 5μg/mL hemin, 1μg/mL menadione (Sigma Aldrich, St. Louis, MO), at 37°C in an anaerobic cabinet. *Streptococcus* strains were plated on TSB with 1.5% w/v agar (Neogen, Lansing, MI) supplemented with 5% defibrinated sheep blood (Northeast Laboratory, Waterville, ME), and grown overnight in TSB at 37°C in 5% CO_2. Colony forming units (CFU) were measured by serial dilution and plating.

Mice

All mice were maintained in a specific pathogen free barrier facility at the Harvard School of Public Health (HSPH) and all experimentation was carried out in accordance with institutional guidelines. Female $Apc^{Min/+}$ mice were ordered from Jackson Laboratory (Jackson strain C57BL/6J-Apc^{Min}/J) and received by 5 weeks of age and $Apc^{Min/+}$ were bred at HSPH. BALB/c *Il-10*$^{-/-}$ and *T-bet*$^{-/-}$ *Rag2*$^{-/-}$ mice were maintained and bred within the HSPH barrier facility.

Bacterial feeding experiments were performed for a period of 8 weeks, beginning at 6 weeks of age. Bacteria were fed to mice by pipette at 10^8 CFU per day, at a maximum volume of 100μL. Sham treatment consisted of 100μL of TSB.

Histopathology

Colons and small intestines were prepared for histologic analysis and assessed for colitis as previously described [49]. Adenomas (polyps with low-grade or high-grade epithelial dysplasia but no invasion) and invasive adenocarcinomas were counted in colons and small intestines submitted in their entirety.

Human specimen collection

Colonic adenocarcinoma samples: Patients were identified from upcoming operative cases by co-investigator T.C. in the Department of Surgery at Brigham and Women's Hospital. Inclusion criteria were patients who had biopsy-confirmed colorectal cancer and were undergoing hemicolectomy. Exclusion criteria included a known synchronous cancer diagnosis or other cancer diagnosis within five (5) years of the operation. The attending surgeon is a study investigator and obtained written informed consent preoperatively. Once the specimen was removed, a portion of the tumor was allocated to members of the study and the remainder submitted for standard processing by the Pathology Department. No antibiotics were given pre-operatively. Written informed consent was obtained from all participants.

Colonic adenoma samples: Patients were identified and samples were collected by co-investigator P.L. (U. of Aberdeen), D.C. (Massachusetts General Hospital), or by the Cooperative Human Tissue Network (U. of Pennsylvania, U. of Virginia Health System, Vanderbilt U. Medical Center). Written informed consent was obtained from all participants.

Stool samples: Samples were collected by the National Cancer Institute Early Detection Research Network. Written informed consent was obtained from each study participant. Colonic adenocarcinoma was confirmed by biopsy; colonic adenoma was confirmed by colonoscopy. All samples were collected pre-operatively. The samples were collected by subjects in their homes

using a “hat” or specimen collection container that fit under a toilet seat, then stool was packed into glass vials and placed on ice packs until shipped overnight and then stored at -80°C.

DNA preparation and bacterial quantification by qPCR

Human and mouse tissues or stools (20-100mg) were digested overnight in 0.7mL molecular grade lysis buffer (100mM TrisHCl pH 8.5, 5mM EDTA pH 8.0, 0.2% SDS, 200mM NaCl, 1 mg/ml proteinase K) at 55°C with rotation. The samples were centrifuged at 20,000x*g* for 5min then the liquid portion was moved to equal volume isopropanol. The precipitated DNA was recovered and resuspended in 0.4mL TE buffer. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added, mixed by inversion, centrifuged at 20,000x*g* for 5min, then the aqueous phase was transferred to a fresh tube. If aqueous phase was milky, the phenol:chloroform:isoamyl alcohol step was repeated. 0.1 volume of 3M sodium acetate pH 5.2 and 2 volumes of 100% ethanol was added, incubated at -20C for 1hr, and centrifuged at 20,000x*g* for 15min. The pellet was washed 2 times by adding 0.5mL of 70% ethanol, centrifuging at 20,000x*g* for 7min, and discarding supernatant. The DNA pellet was resuspended in molecular grade water. For stool samples, an extra purification step was performed at this stage using the QIAquick Gel Extraction Kit (QIAGEN).

8ng of DNA was used in each 20µL KAPA SYBR® FAST qPCR (Kapa Biosystems, Woburn, MA) reaction, performed in triplicate, and analyzed on the Stratagene Mx3005P (Agilent Technologies, Santa Clara, CA). The following primer sets were used: *Fusobacterium* spp. (Fwd 5’-GGATTTATTGGGCGTAAAGC-3’; Rev 5’-GGCATTCCTACAAATATCTACGAA-3’) [50], and universal Eubacteria 16S (Fwd 5’-GGTGAATACGTTCCCGG-3’; Rev 5’-TACGGCTACCTTGTTACGACTT-3’). Relative abundance was calculated by the ΔC_T method.

RNA preparation and gene expression by qPCR

Total RNA was extracted using the RNeasy® kit (QIAGEN), DNase treated with the DNA-*free*™ kit (Ambion), and cDNA was generated using the iScript™ cDNA Synthesis kit (Bio-Rad). All primers were ordered from Sigma Aldrich. Relative gene expression was calculated using the $\Delta\Delta C_T$ method. Primers were designed using MGH primerbank: *pga.mgh.harvard.edu/primerbank/*

Microbial FISH analysis

Microbial FISH was performed as described previously [7].

RNA-Seq processing and analysis

All primary sequence data on colonic adenoma (COAD) samples (https://tcga-data.nci.nih.gov/docs/publications/coadread_2012/) were downloaded from The Cancer Genome Atlas via dbGAP and the Data Coordinating Center (http://cancergenome.nih.gov/). At the time of analysis, 133 complete COAD RNA-Seq datasets were available. All datasets were analyzed on PathSeq [29] to calculate microbial relative abundance at the genus level. Relative abundance was defined as the number of uniquely-mapped microbial reads to a given genus, normalized for the total number of uniquely-mapped microbial reads per sample. Host gene expression was calculated using RPKM as described previously [28]. The Spearman's rank correlation coefficient was calculated for each genus-host gene pairing.

Detection of NF-κB activation

Fresh or fresh-frozen colon tumor tissue (20-80mg per sample) was processed for nuclear protein isolation with the NE-PER® Nuclear and Cytoplasmic Reagents Kit (Thermo Scientific, Rockford, IL). Protein lysates were resolved using SDS-PAGE and transferred to PVDF membrane using a Bio-Rad wet transfer apparatus. Blots were probed with antibodies directed

against p65 and lamin B1 (Cell Signaling Technology, Danvers, MA). Samples were stratified as *Fusobacterium*-high (qPCR *Fusobacterium* abundance greater than 40.0 [calculated as 1.8^(-C_T<*Fusobacterium*> - C_T<Eubacteria>)*1000], or 16S sequencing *Fusobacterium* abundance greater than 0.25 [calculated as described previously [7]]) or *Fusobacterium*-low (qPCR *Fusobacterium* abundance less than 1.0, or 16S sequencing *Fusobacterium* abundance less than 0.015).

Isolation of intestinal tumor infiltrating cells

Small intestine was removed, cut longitudinally and washed with Ca^+- and Mg^+-free Dulbecco's phosphate buffered saline (DPBS). Tumor nodules were dissected away from normal tissue, weighed, and then finely minced and incubated in Ca^+- and Mg^+-free Hank's balanced salt solution (HBSS) with 0.1 mg/ml collagenase D (Roche) and 50 U/ml DNase I (Roche) for 30 min at 37°C on a shaking platform. The solution containing digested tumors was filtered through a 70 μm cell strainer and centrifuged at 400 X *g* for 10 min. Isolated tumor infiltrating cells were resuspended in cell staining solution (PBS with 2% FCS) for flow cytometry analysis.

Flow cytometry

For cell surface staining, cells were incubated with Fc blocking antibody (BioLegend) for 15 min, and stained with fluorochrome-conjugated monoclonal antibodies of cell surface markers. Following antibodies were used and purchased from BioLegend or eBioscience: CD45 (clone 30-F11), CD11b (clone M1/70), Gr-1 (clone RB6-8C5), CD11c (clone N418), MHC class II (clone M5/114.15.2), CD3ε (clone 145-2C11), CD4 (clone RM4.5), CD8α (clone 53-6.7), F4/80 (clone BM8), Ly6C (cloneHK1.4), CD103 (clone 2E7), and CD16/CD32 (clone CD16/CD32). For intracellular cytokine or Foxp3 staining, cells were stimulated for 4 h in RPMI complete medium with 50 ng/mL phorbol 12-myristate 13-acetate (PMA)(Sigma) and 500

ng/mL ionomycin (Sigma) in the presence of 5ug/mL Brefeldin A (BioLegend). Then, cells were stained with cell surface markers, fixed with Fix/Perm buffer (BioLegend), permeabilized with Perm buffer (BioLegend) according to the manufacturer's recommendations and stained with anti-IL-17A (clone TC11-18H10.1, BioLegend) or anti-Foxp3 (clone FJK-16s, eBioscience) antibodies. Cells were stained in parallel with the respective control isotype antibodies. Stained cells were acquired using BD LSRII flow cytometry (BD Biosciences) and analyzed with FlowJo9 software (Tree Star).

Statistical analysis

Generally, data is displayed in dot-plot format, with the center-line indicating the mean and the standard error of the mean represented by the error bars. All two-group comparisons were performed using the non-parametric Mann-Whitney U test, with the exception of **Fig. 3-4a** for which the non-parametric Wilcoxon matched-pairs signed rank test was performed. For RNA-Seq data, genus relative abundance was correlated to host gene expression using the Spearman's rank correlation coefficient, and corrected *P*-values were obtained by correcting for multiple hypothesis testing using the False Discovery Rate method. Statistical analysis for **Supplementary Table 3-2** was performed using the Fisher's exact test.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D: **Global cancer statistics.** *CA Cancer J Clin* 2011, **61**:69–90.

2. Kado S, Uchida K, Funabashi H, Iwata S, Nagata Y, Ando M, Onoue M, Matsuoka Y, Ohwaki M, Morotomi M: **Intestinal microflora are necessary for development of spontaneous adenocarcinoma of the large intestine in T-cell receptor beta chain and p53 double-knockout mice. [Internet]**. *Cancer Res.* 2001, **61**:2395–2398.

3. Garrett WS, Punit S, Gallini CA, Michaud M, Zhang D, Sigrist KS, Lord GM, Glickman JN, Glimcher LH: **Colitis-associated colorectal cancer driven by T-bet deficiency in dendritic cells.** *Cancer Cell* 2009, **16**:208–219.

4. Uronis JM, Mühlbauer M, Herfarth HH, Rubinas TC, Jones GS, Jobin C: **Modulation of the intestinal microbiota alters colitis-associated colorectal cancer susceptibility.** *PLoS ONE* 2009, **4**:e6026.

5. Wu S, Rhee K-J, Albesiano E, Rabizadeh S, Wu X, Yen H-R, Huso DL, Brancati FL, Wick E, McAllister F, et al.: **A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses.** *Nat. Med.* 2009, **15**:1016–1022.

6. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan T-J, Campbell BJ, Abujamel T, Dogan B, Rogers AB, et al.: **Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota**. *Science* 2012, **338**:120–123.

7. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, Ojesina AI, Jung J, Bass AJ, Tabernero J, et al.: **Genomic analysis identifies association of Fusobacterium with colorectal carcinoma**. *Genome Res* 2012, **22**:292–298.

8. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, Barnes R, Watson P, Allen-Vercoe E, Moore RA, et al.: **Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma**. *Genome Res* 2012, **22**:299–306.

9. Marchesi JR, Dutilh BE, Hall N, Peters WHM, Roelofs R, Boleij A, Tjalsma H: **Towards the Human Colorectal Cancer Microbiome**. *PLoS ONE* 2011, **6**:e20447.

10. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE: **A bacterial driver–passenger model for colorectal cancer: beyond the usual suspects**. *Nature Reviews Microbiology* 2012, **10**:575–582.

11. Macaluso A, Simmang C, Anthony T: **Streptococcus sanguis Bacteremia and Colorectal Cancer.** *Southern Medical Journal* 1998, **91**:206.

12. Mantovani A, Allavena P, Sica A, Balkwill F: **Cancer-related inflammation**. *Nature* 2008, **454**:436–444.

13. Mantovani A, Sica A: **Macrophages, innate immunity and cancer: balance, tolerance, and diversity.** *Current Opinion in Immunology* 2010, **22**:231–237.

14. Qian B-Z, Pollard JW: **Macrophage diversity enhances tumor progression and metastasis.** *Cell* 2010, **141**:39–51.

15. Ruffell B, Affara NI, Coussens LM: **Differential macrophage programming in the tumor microenvironment.** *Trends Immunol.* 2012, **33**:119–126.

16. Gabrilovich DI, Castellarin M, Nagaraj S, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, Barnes R, Watson P, et al.: **Myeloid-derived suppressor cells as regulators of the immune system.** *Nat. Rev. Immunol.* 2009, **9**:162–174.

17. Ostrand-Rosenberg S: **Myeloid-derived suppressor cells: more mechanisms for inhibiting antitumor immunity.** *Cancer Immunol. Immunother.* 2010, **59**:1593–1600.

18. Ostrand-Rosenberg S, Sinha P, Beury DW, Clements VK: **Cross-talk between myeloid-derived suppressor cells (MDSC), macrophages, and dendritic cells enhances tumor-induced immune suppression.** *Semin. Cancer Biol.* 2012, **22**:275–281.

19. Gabrilovich DI, Arthur JC, Ostrand-Rosenberg S, Perez-Chanona E, Bronte V, Mühlbauer M, Tomkovich S,

Uronis JM, Fan T-J, Campbell BJ, et al.: **Coordinated regulation of myeloid cells by tumours.** *Nat. Rev. Immunol.* 2012, **12**:253–268.

20. Mellman I, Coukos G, Dranoff G: **Cancer immunotherapy comes of age.** *Nature* 2011, **480**:480–489.

21. Apetoh L, Locher C, Ghiringhelli F, Kroemer G, Zitvogel L: **Harnessing dendritic cells in cancer.** *Semin. Immunol.* 2011, **23**:42–49.

22. Coombes JL, Garrett W, Siddiqui KRR, Arancibia-Cárcamo CV, Hall J, Sun C-M, Belkaid Y, Powrie F: **A functionally specialized population of mucosal CD103+ DCs induces Foxp3+ regulatory T cells via a TGF-beta and retinoic acid-dependent mechanism.** *J. Exp. Med.* 2007, **204**:1757–1764.

23. Siddiqui KRR, Ohkusa T, Wu S, Laffont S, Yoshida T, Rhee K-J, Powrie F, Sato N, Albesiano E, Watanabe S, et al.: **E-cadherin marks a subset of inflammatory dendritic cells that promote T cell-mediated colitis.** *Immunity* 2010, **32**:557–567.

24. Josefowicz SZ, Lu L-F, Rudensky AY: **Regulatory T cells: mechanisms of differentiation and function.** *Annu. Rev. Immunol.* 2012, **30**:531–564.

25. Grivennikov SI, Wang K, Mucida D, Stewart CA, Schnabl B, Jauch D, Taniguchi K, Yu G-Y, Österreicher CH, Hung KE, et al.: **Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth**. *Nature* 2012, doi:10.1038/nature11465.

26. Yu H, Kortylewski M, Pardoll D: **Crosstalk between cancer and immune cells: role of STAT3 in the tumour microenvironment.** *Nat. Rev. Immunol.* 2007, **7**:41–51.

27. Ostrand-Rosenberg S, Sinha P: **Myeloid-Derived Suppressor Cells: Linking Inflammation and Cancer**. *J. Immunol.* 2009, **182**:4499–4506.

28. Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, Kovar CL, Lewis LR, Morgan MB, Newsham IF, et al.: **Comprehensive molecular characterization of human colon and rectal cancer**. *Nature* 2012, **487**:330–337.

29. Kostic AD, Ojesina AI, Pedamallu CS, Jung J, Verhaak RGW, Getz G, Meyerson M: **PathSeq: software to identify or discover microbes by deep sequencing of human tissue.** *Nat. Biotechnol.* 2011, **29**:393–396.

30. Dharmani P, Strauss J, Ambrose C, AllenVercoe E, Chadee K: **Fusobacterium nucleatum infection of colonic cells stimulates MUC2 mucin and tumor necrosis factor-{alpha}**. *Infection and Immunity* 2011, doi:10.1128/IAI.05118-11.

31. Grivennikov SI, Karin M: **Inflammatory cytokines in cancer: tumour necrosis factor and interleukin 6 take the stage.** *Ann. Rheum. Dis.* 2011, **70 Suppl 1**:i104–8.

32. Trinchieri G: **Cancer and inflammation: an old intuition with rapidly evolving new concepts.** *Annu. Rev. Immunol.* 2012, **30**:677–706.

33. Hayden MS, Hayden MS, Ghosh S, Ghosh S: **NF-κB, the first quarter-century: remarkable progress and outstanding questions.** *Genes & Development* 2012, **26**:203–234.

34. DiDonato JA, Ruffell B, Saleh M, Mantovani A, Segata N, Mercurio F, Affara NI, Trinchieri G, Allavena P, Haake SK, et al.: **NF-κB and the link between inflammation and cancer.** *Immunol. Rev.* 2012, **246**:379–400.

35. Scanlan PD, Shanahan F, Clune Y, Collins JK, O'Sullivan GC, O'Riordan M, Holmes E, Wang Y, Marchesi JR: **Culture-independent analysis of the gut microbiota in colorectal cancer and polyposis.** *Environ.*

Microbiol. 2008, **10**:789–798.

36. Shen XJ, Rawls JF, Randall T, Burcal L, Mpande CN, Jenkins N, Jovov B, Abdo Z, Sandler RS, Keku TO: **Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas.** *Gut Microbes* 2010, **1**:138–147.

37. Sanapareddy N, Legge RM, Jovov B, McCoy A, Burcal L, Araujo-Perez F, Randall TA, Galanko J, Benson A, Sandler RS, et al.: **Increased rectal microbial richness is associated with the presence of colorectal adenomas in humans.** *ISME J* 2012, **6**:1858–1868.

38. McCoy AN, Araujo-Perez F, Azcárate-Peril A, Yeh JJ, Sandler RS, Keku TO: **Fusobacterium Is Associated with Colorectal Adenomas**. *PLoS ONE* 2013, **8**:e53653.

39. Jobin C: **Colorectal cancer: CRC-all about microbial products and barrier function?** *Nat Rev Gastroenterol Hepatol* 2012, **9**:694–696.

40. Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D, Huttenhower C, Izard J: **Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples**. *Genome Biol.* 2012, **13**:R42.

41. Cho KR, Vogelstein B: **Genetic alterations in the adenoma--carcinoma sequence.** *Cancer* 1992, **70**:1727–1731.

42. Han YW, Shi W, Huang GTJ, Kinder Haake S, Park N-H, Kuramitsu H, Genco RJ: **Interactions between Periodontal Bacteria and Human Oral Epithelial Cells: Fusobacterium nucleatum Adheres to and Invades Epithelial Cells**. *Infection and Immunity* 2000, **68**:3140–3146.

43. Strauss J, Kaplan GG, Beck PL, Rioux K, Panaccione R, Devinney R, Lynch T, Allen-Vercoe E: **Invasive potential of gut mucosa-derived Fusobacterium nucleatum positively correlates with IBD status of the host.** *Inflamm. Bowel Dis.* 2011, **17**:1971–1978.

44. Vander Heiden MG, Grivennikov SI, Han YW, Cantley LC, Wang K, Shi W, Thompson CB, Mucida D, Huang GT, Stewart CA, et al.: **Understanding the Warburg effect: the metabolic requirements of cell proliferation.** *Science* 2009, **324**:1029–1033.

45. Kapatral V, Kapatral V, Anderson I, Anderson I, Ivanova N, Ivanova N, Reznik G, Reznik G, Los T, Los T, et al.: **Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586.** *J. Bacteriol.* 2002, **184**:2005–2018.

46. Gursoy UK, Hanahan D, Pöllänen M, Weinberg RA, Könönen E, Uitto V-J: **Biofilm formation enhances the oxygen tolerance and invasiveness of Fusobacterium nucleatum in an oral mucosa culture model.** *J. Periodontol.* 2010, **81**:1084–1091.

47. Chan AT, Arber N, Burn J, Chia WK, Elwood P, Hull MA, Logan RF, Rothwell PM, Schrör K, Baron JA: **Aspirin in the Chemoprevention of Colorectal Neoplasia: An Overview**. *Cancer Prevention Research* 2012, **5**:164–178.

48. Consortium THMP: **Structure, function and diversity of the healthy human microbiome**. *Nature* 2012, **486**:207–214.

49. Garrett WS, Lord GM, Punit S, Lugo-Villarino G, Mazmanian SK, Ito S, Glickman JN, Glimcher LH: **Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system.** *Cell* 2007, **131**:33–45.

50. Boutaga K, van Winkelhoff AJ, Vandenbroucke-Grauls CMJE, Savelkoul PHM: **Periodontal pathogens: A**

quantitative comparison of anaerobic culture and real-time PCR. *FEMS Immunol. Med. Microbiol.* 2005, **45**:191–199.

This page intentionally left blank

CONCLUSION

Perspectives on the Study of the Microbiota in Colorectal Cancer

This thesis describes the composition of the colorectal tumor microbiome and explores the functional consequences of some of its constituents to tumorigenesis. First, we describe the development of new computational tools that were used in this analysis. PathSeq builds on the conceptual framework of computational subtraction, and for the first time, makes it technically feasible to analyze human whole-genome sequence datasets for the purpose of pathogen discovery. Second, we make use of whole-genome sequencing using PathSeq to perform the first large-scale study of the colorectal cancer microbiome. In performing this analysis, we identified a striking enrichment of the *Fusobacterium* genus in tumors. Although *Fusobacterium* is a rare gut constituent in healthy individuals, almost every colon tumor sequenced was positive for *Fusobacterium*. We then introduced an invasive human gut isolate of *Fusobacterium nucleatum* into $Apc^{Min/+}$ mice and observed accelerated intestinal tumorigenesis as well as a tumor-infiltrating myeloid cell inflammatory signature.

The value of these findings remains to be determined. An inherent limitation of any microbiome study that compares the composition of healthy versus diseased tissues or subjects is that the conclusions are correlative; such studies cannot address causation. A long-term prospective cohort study can be more informative than a case-control study because it allows for the identification of a particular shift in microbial composition at the time of disease onset. However, such a study design is significantly more complex, and it still cannot reveal whether the shift in microbial composition has a causal component or is merely a consequence of the disease state. An experimental animal-based study such as the one we have performed does address causation, however there are significant limitations in extrapolating results from animal studies to human disease. Therefore, taking the case of *Helicobacter pylori* and gastric cancer as an example, to assess whether *Fusobacterium* species can have a causal role in human colorectal

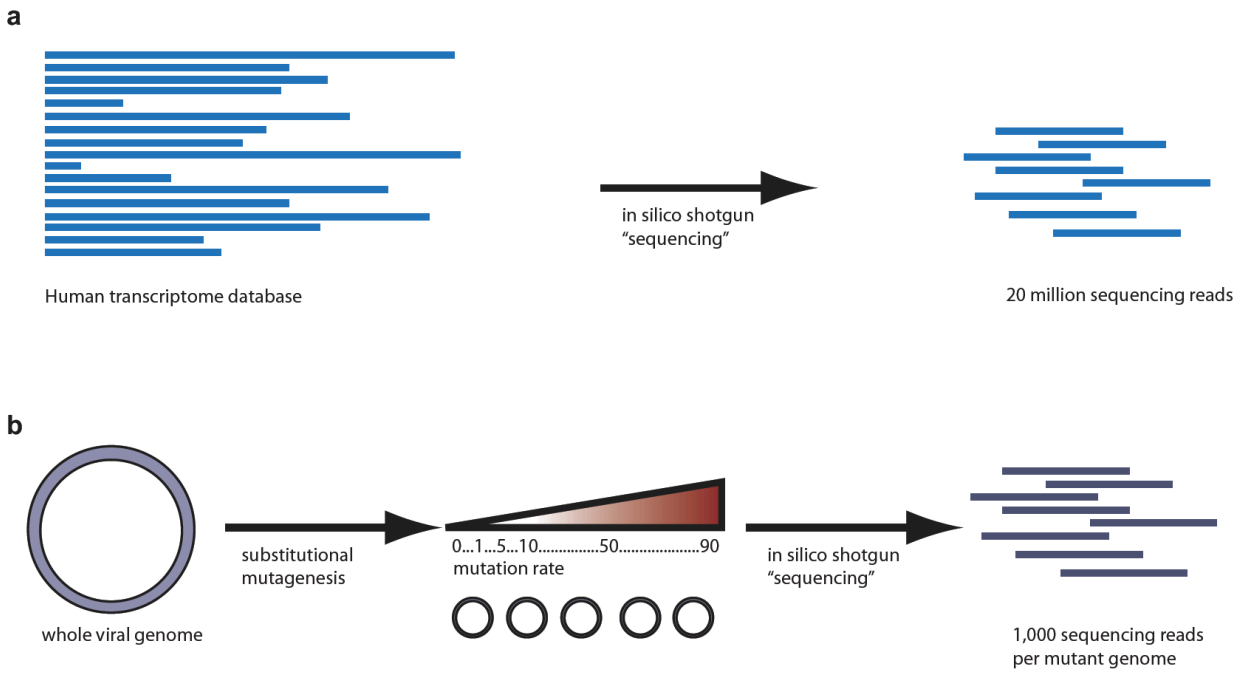
cancer we must combine results from epidemiological and case-control studies with mechanistic animal-based studies and build evidence in a cumulative fashion.

If *Fusobacterium* species do contribute to human colorectal tumorigenesis then this could have significant clinical implications. An exciting aspect of understanding the role of the microbiota in disease is that the microbiome, unlike the human genome, is extremely malleable. For example, the development and application of narrow-spectrum antibiotics that target the Fusobacteria may prevent or delay the onset of colon cancer in patients that are at high risk of developing the disease. Furthermore, although this route will require significantly more investigation, it is possible that altering the composition of the microbiota with the use specific probiotic formulations and even fecal transplants could deplete *Fusobacterium* species from the gut microbiota.

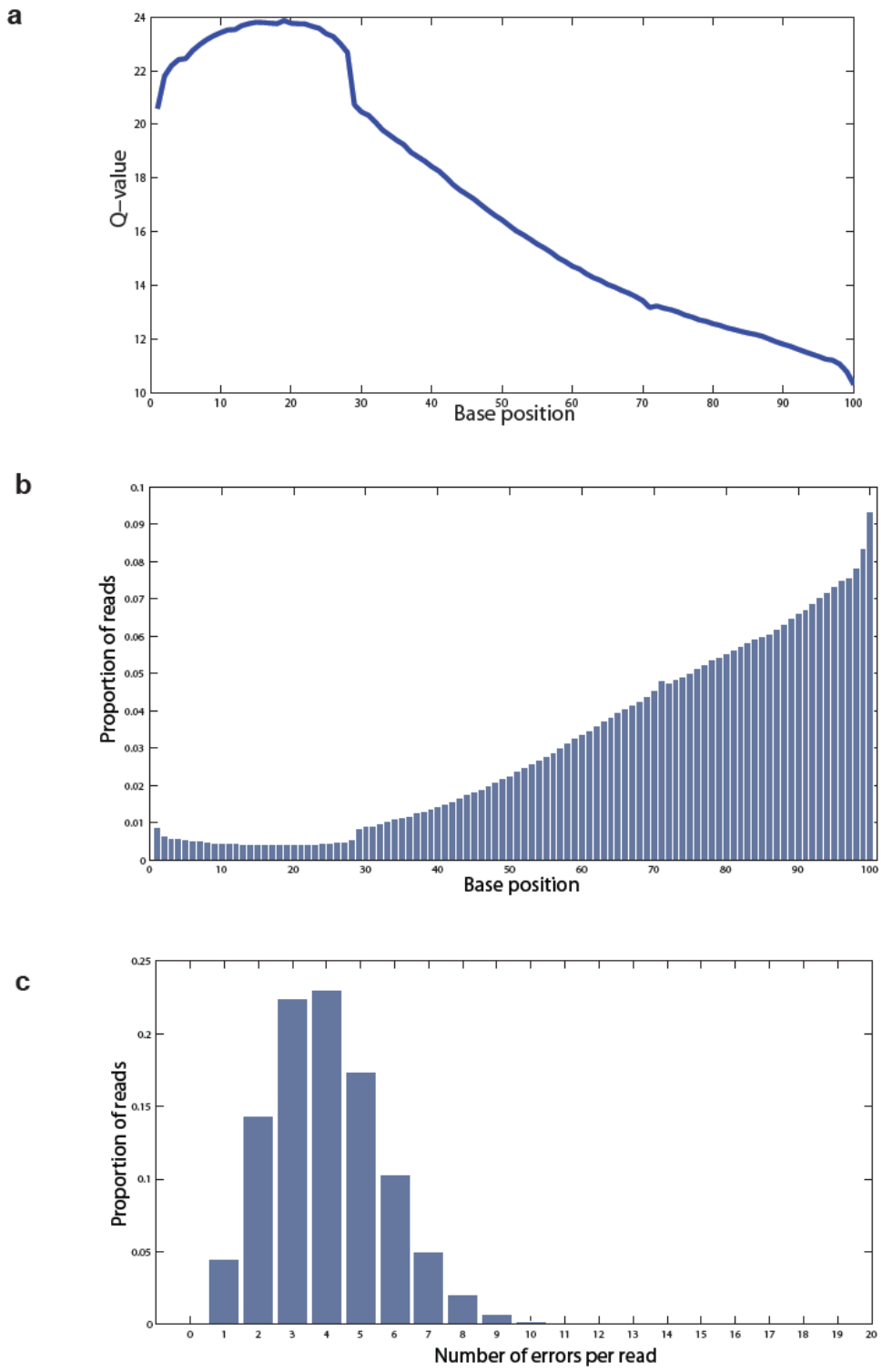
The enrichment of *Fusobacterium* species in colorectal tumors, regardless of its functional contribution to tumorigenesis, may serve as a clinically relevant biomarker. We have demonstrated that *Fusobacterium* spp. can be detected higher levels in stool from patients with colorectal carcinoma, and also in patients with pre-malignant polyps, relative to healthy control individuals. Therefore, such an assay may be of use as a non-invasive clinical diagnostic that could lead to the early detection of disease in some patients.

APPENDIX 1

Supplementary Materials

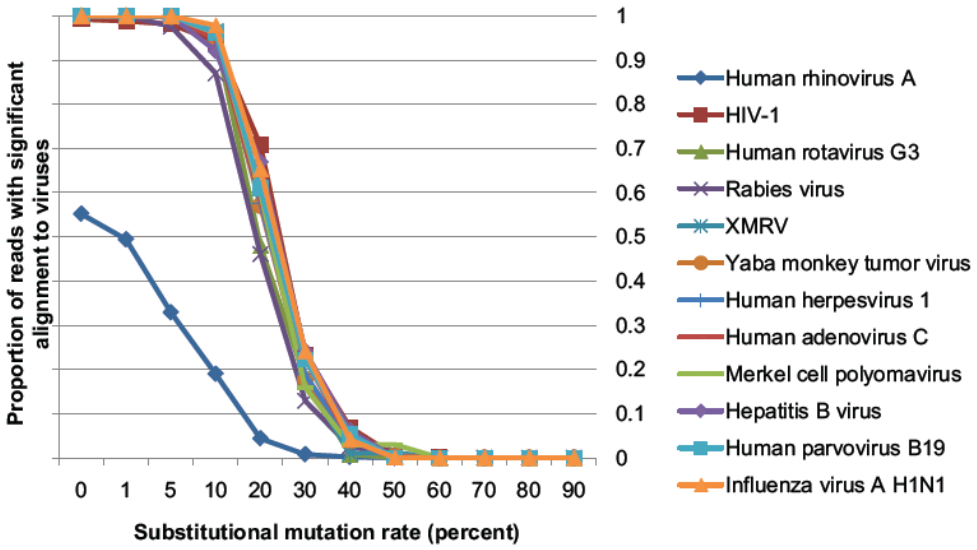


Supplementary Figure 1-1. Generation of artificial shotgun sequencing reads. (**a**) 20 million reads were generated from a human transcriptome database by randomly selecting 100-mer sequences. (**b**) A set of twelve virus genomes were selected. For each genome, substitutional mutations were introduced at twelve distinct mutation rates, and for each of these mutated genomes 100-mer sequences were chosen at random to produce the read set.

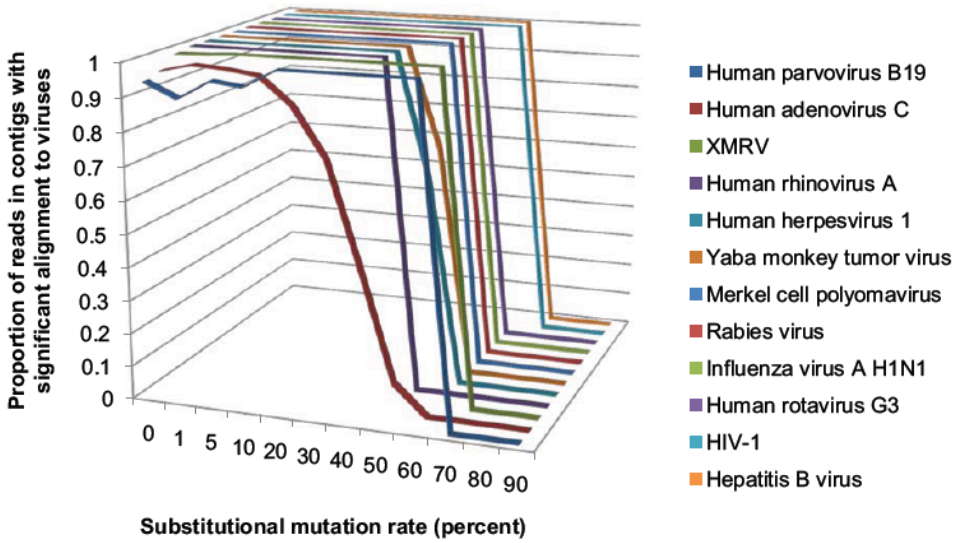


Supplementary Figure 1-2. Applying a sequencing error distribution model to artificially generated reads. (**a**) Average quality score plot for a typical set of reads generated by the Illumina GAII Sequencer. The Q-value is defined as Q=-10*log10(p/(1-p)), where p is the probability that the corresponding base call is incorrect. (**b**) The error distribution was applied to the complete set of reads generated as shown in **Supplementary Figure 1**. The plot shows the proportion of the read set containing a substituted base at the indicated position. (**c**) The proportion of the read set is shown as a function of the number of sequencing errors per read.

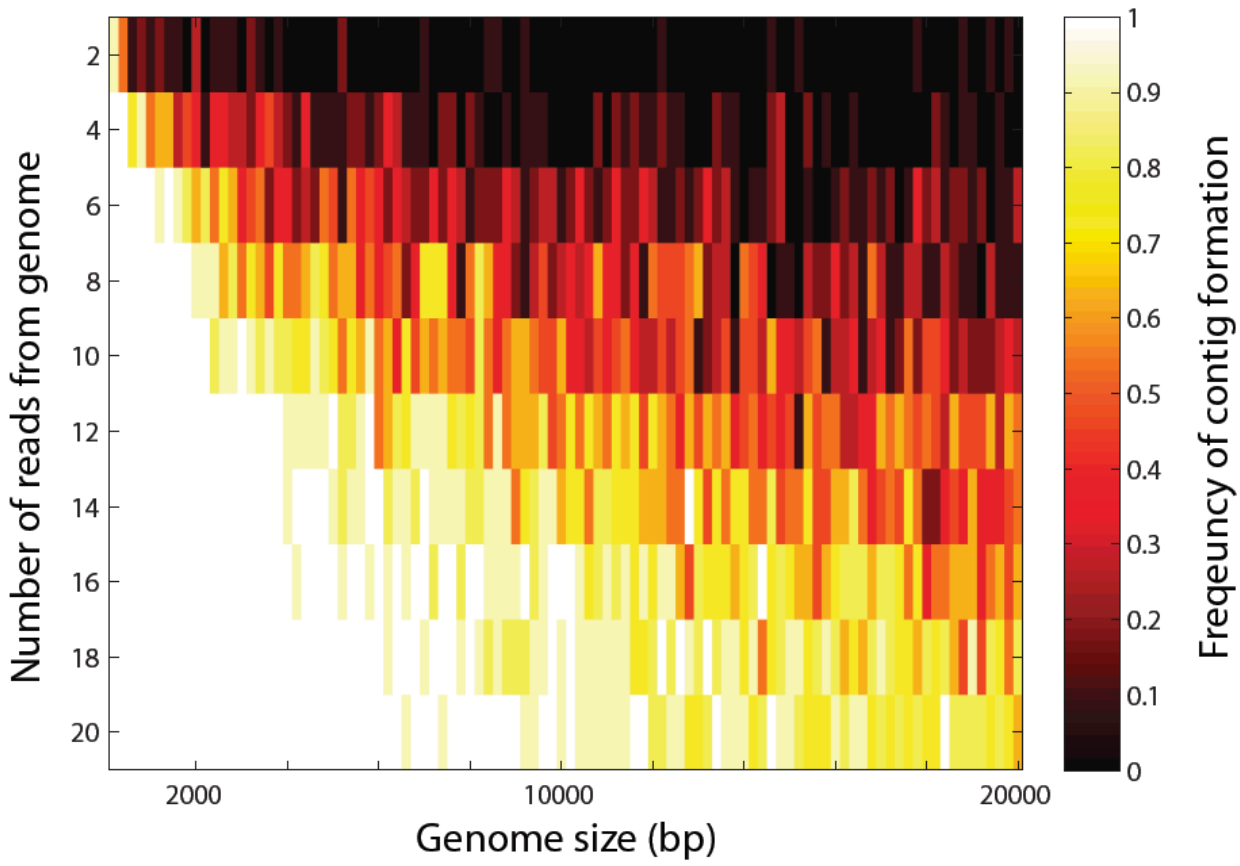
a



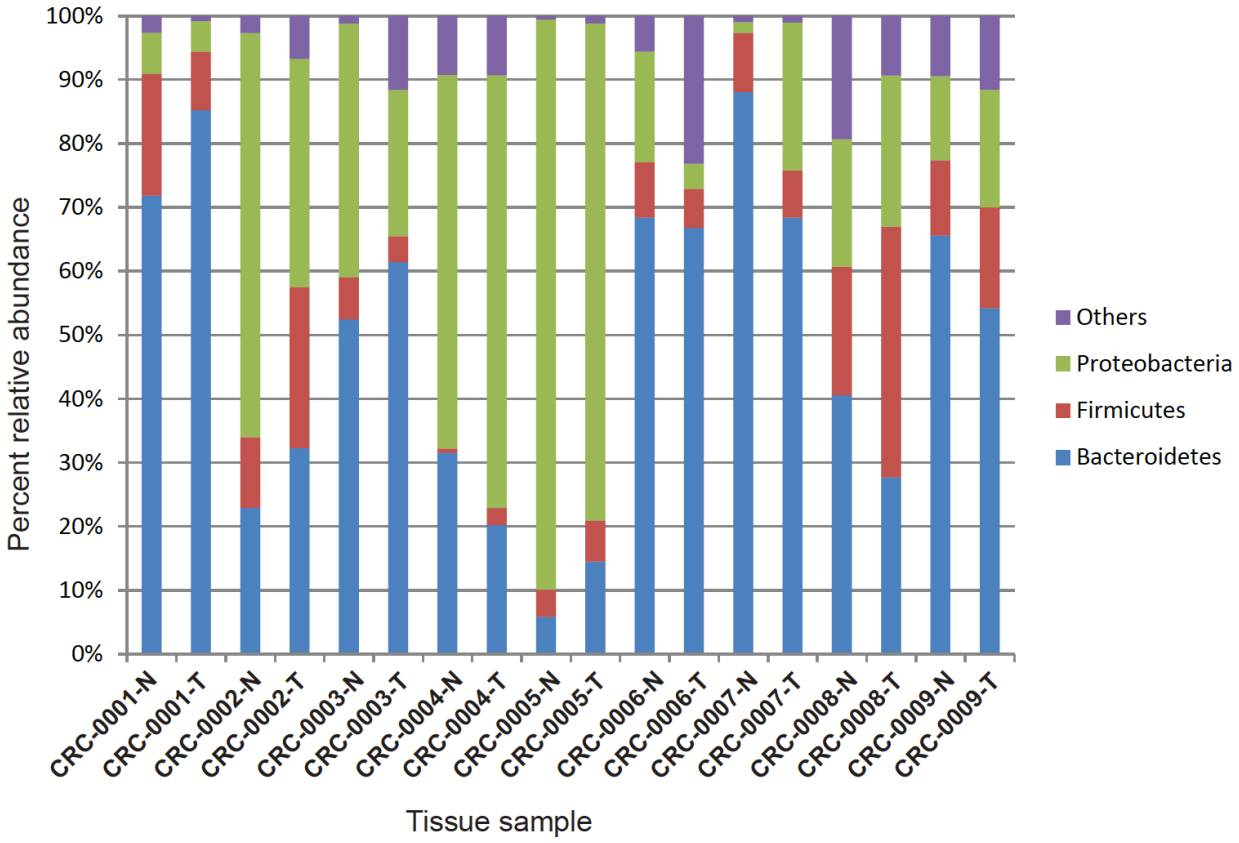
b



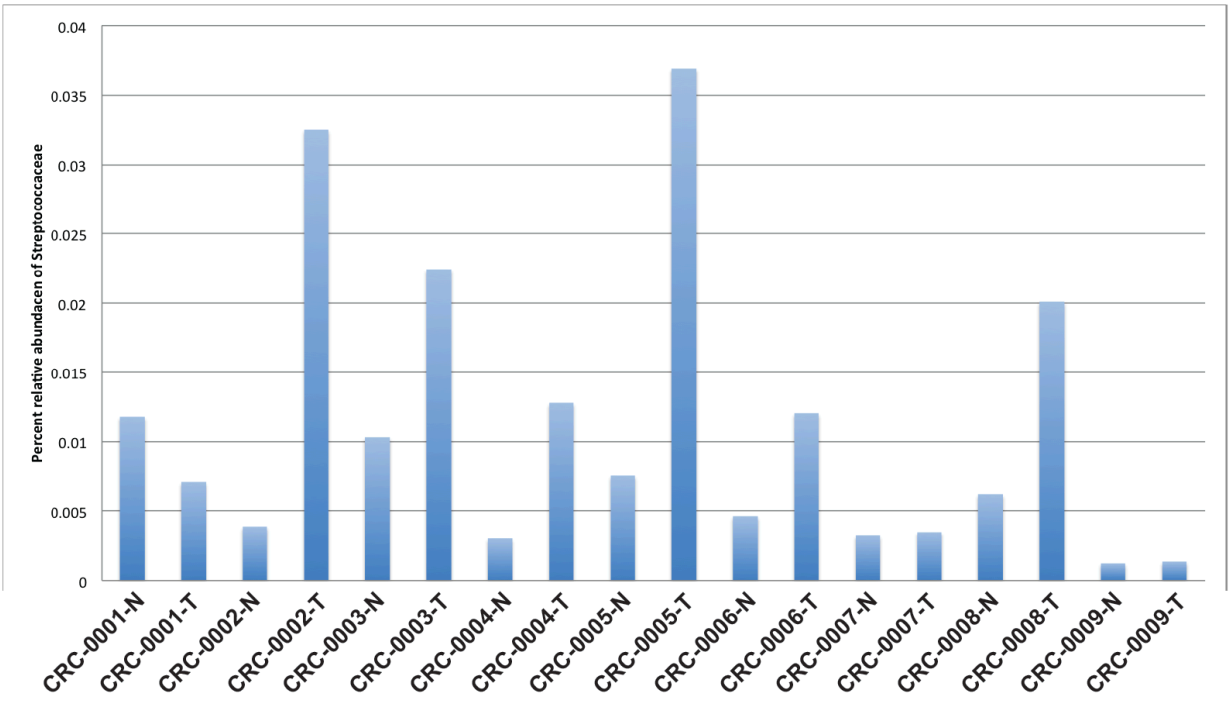
Supplementary Figure 1-3. Identification of virus-derived reads by sequence similarity. (**a**) Reads were generated as indicated in **Supplementary Figure 1-1b**. The proportion of reads identified as a virus sequence (MegaBlast alignment, E-value < 10e-10) is shown as a function of the substitutional mutation rate of the mutated genome. (**b**) Independent assemblies were performed on all reads from each of the 144 genomes. The proportion of reads incorporated into contigs that were identified as a virus sequence (MegaBlast alignment, E-value < 10e-10) is shown.



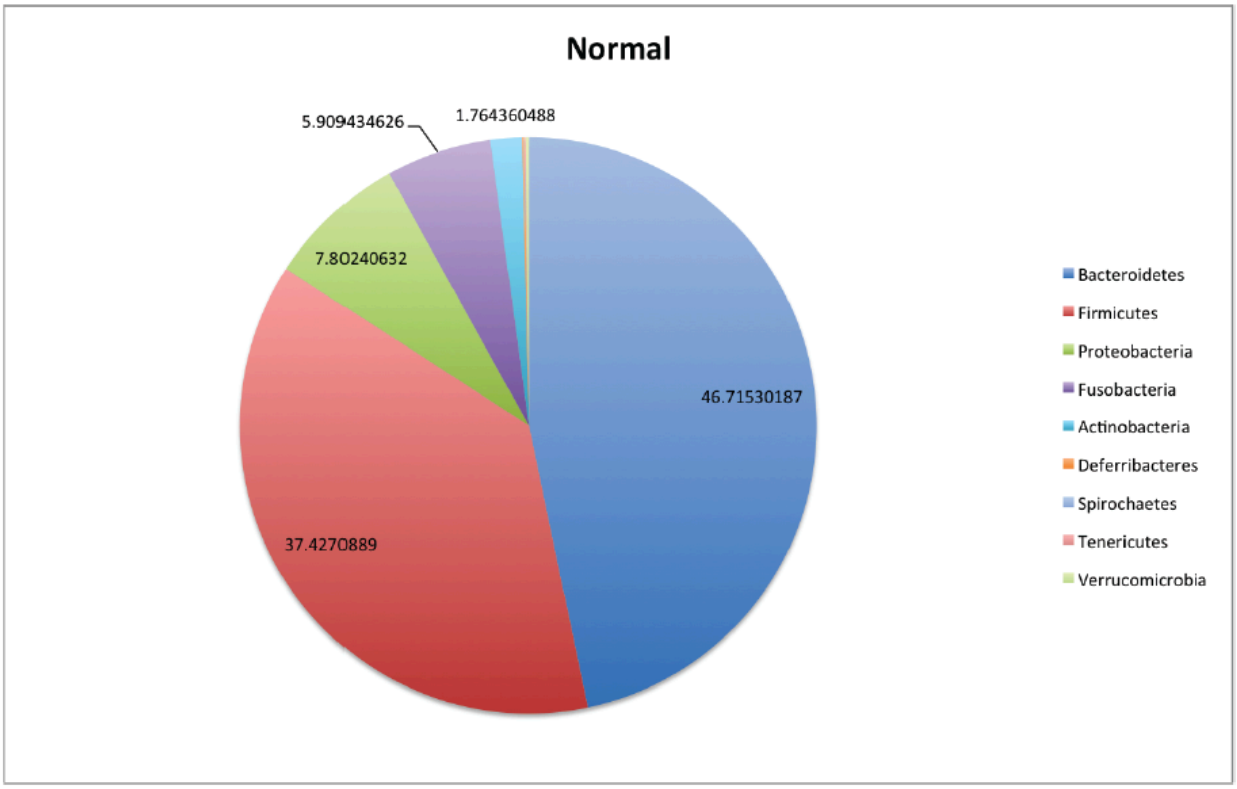
Supplementary Figure 1-4. Probability of contig formation for sequences randomly generated from a genome of varying size. This figure shows the probability of forming contigs from a set of simulated, virus-derived reads. Reads were generated by selecting random 100-mers from a series of genomes ranging in size from 100bp to 20,000bp, producing between 2 and 20 reads per genome. The heatmap indicates the frequency among 11 replicates with which a contig of at least 175bp was formed from these reads by *de novo* assembly.

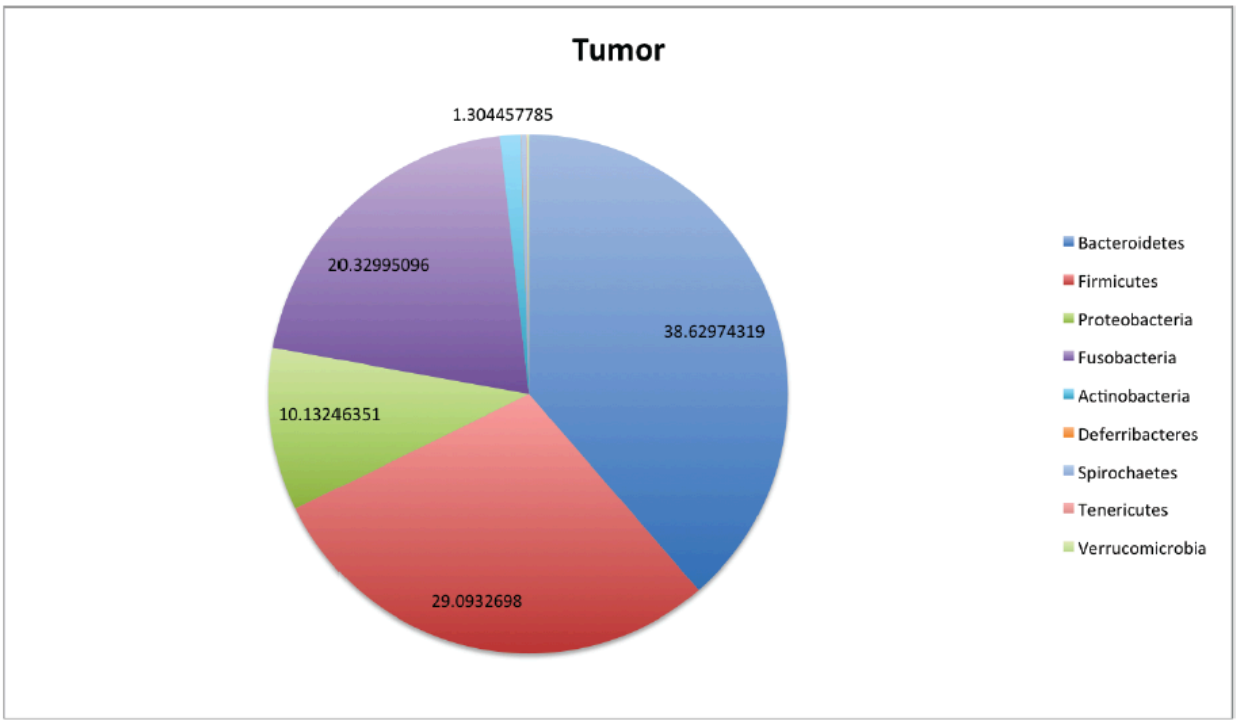


Supplementary Figure 2-1. Phylum-level microbial classifications of whole-genome sequencing on nine colon tumor and normal tissue pairs. Phylum-level classification demonstrates that most specimens are dominated by the Bacteroidetes and Proteobacteria, collectively making up greater than 80% of all phyla in some samples.

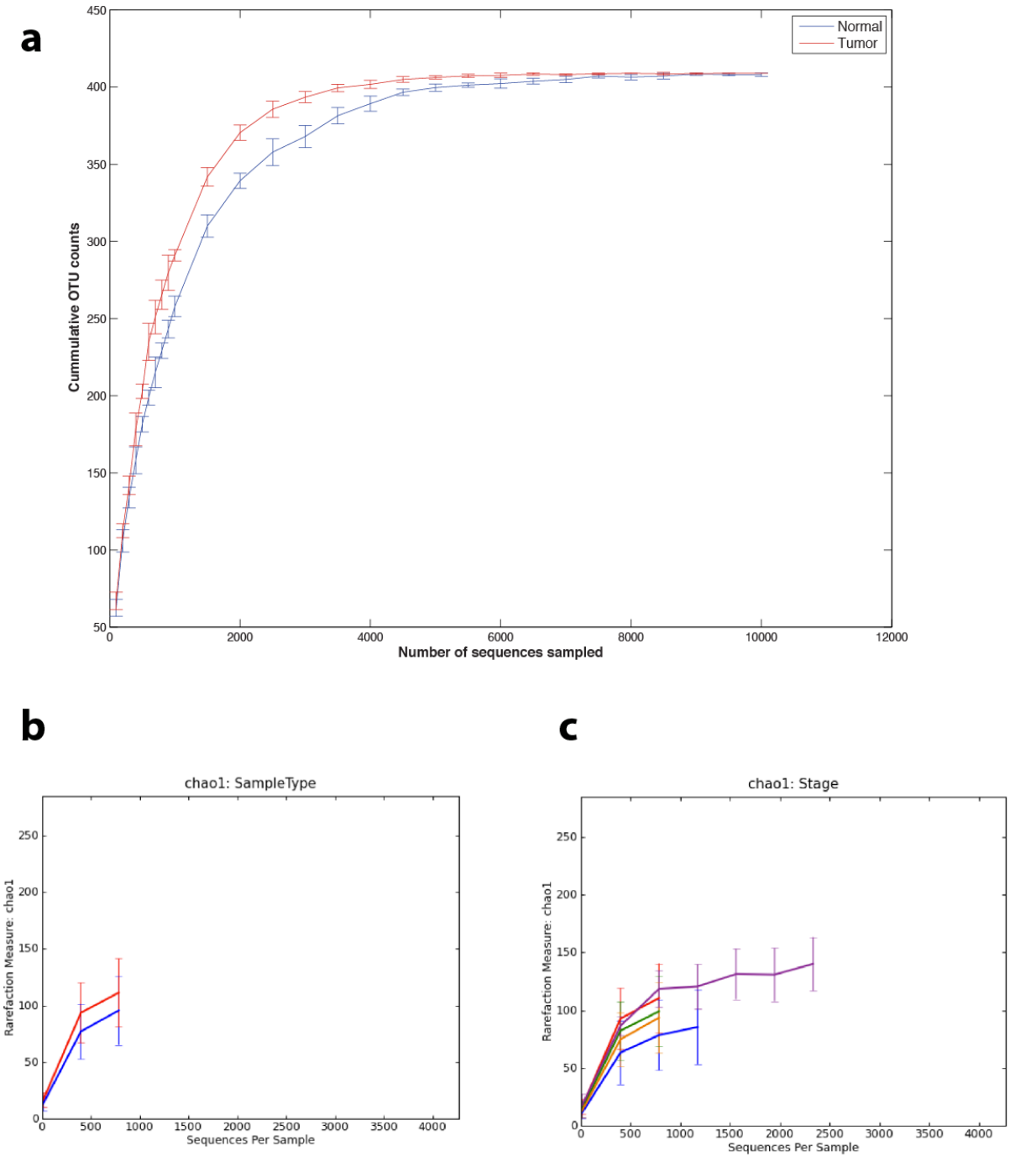


Supplementary Figure 2-2. Whole genome sequencing identifies the *Streptococcaceae* family as significantly enriched in colon tumors. The relative abundance of sequences classified in the *Streptococcaceae* family is shown for nine colon tumor/normal paired samples by whole-genome sequencing. The *Streptococcaceae* family was found to be significantly enriched in tumors. Compare with **Fig. 2-1d** which shows the relative abundance of *Fusobacterium* sequences in the same set of samples.

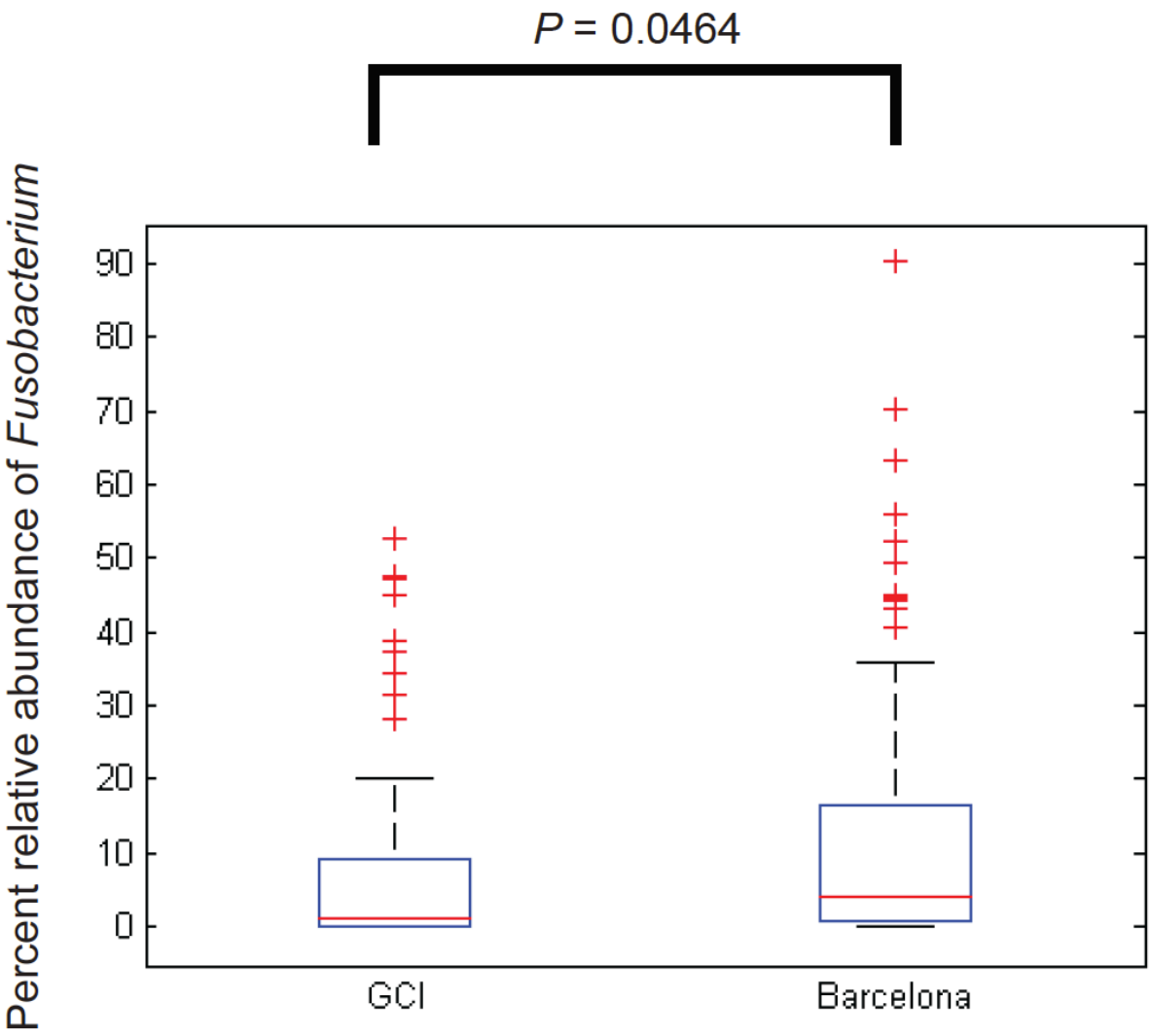




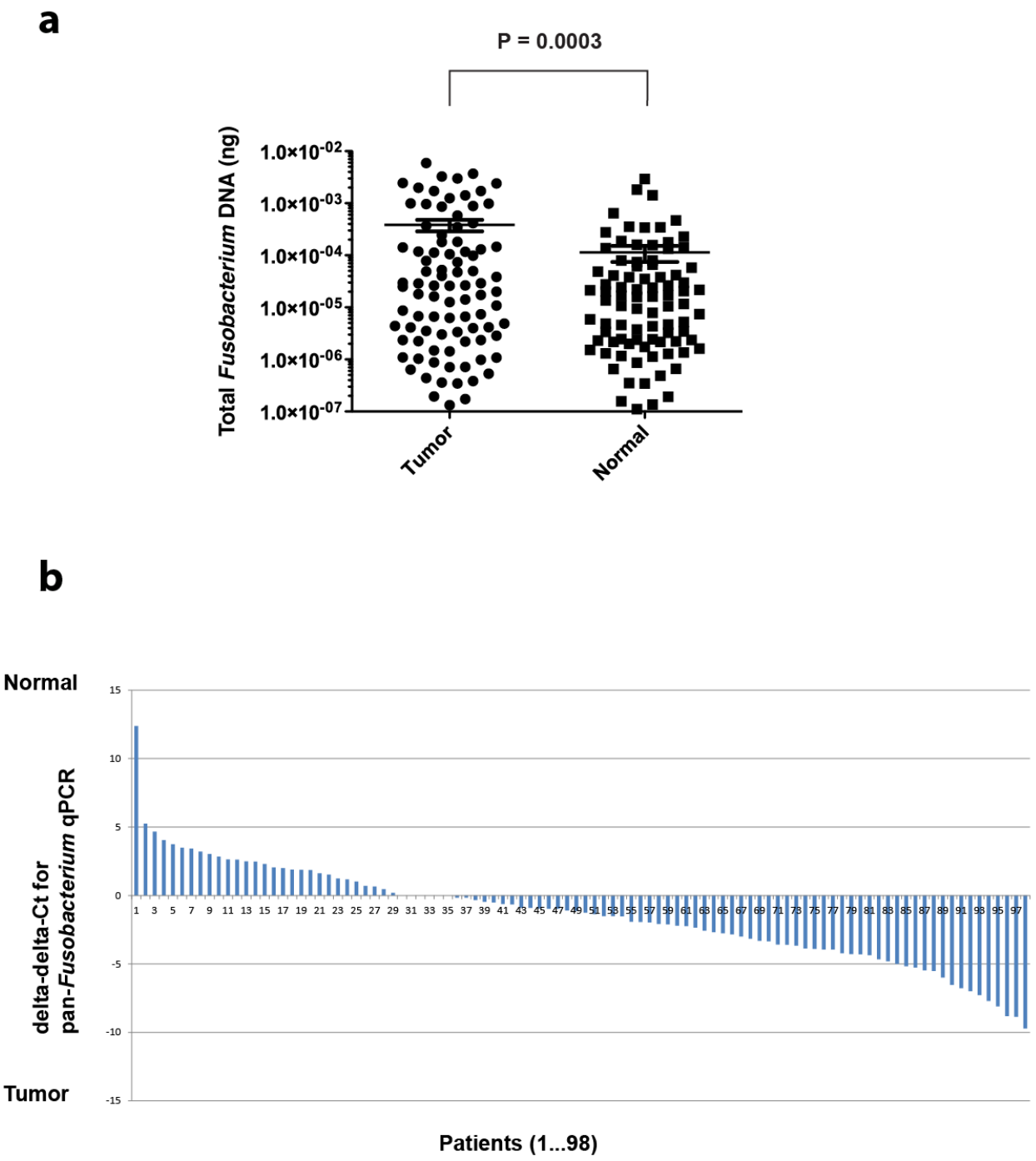
Supplementary Figure 2-3. Phylum-level relative abundance of the microbiota by 16S rDNA sequencing on 95 colon tumor and normal tissue pairs. Median relative abundance measurements at the phylum level are reported across all 95 tumor and normal sample pairs, demonstrating a relative depletion of the Bacteroidetes and Firmicutes in the tumor.



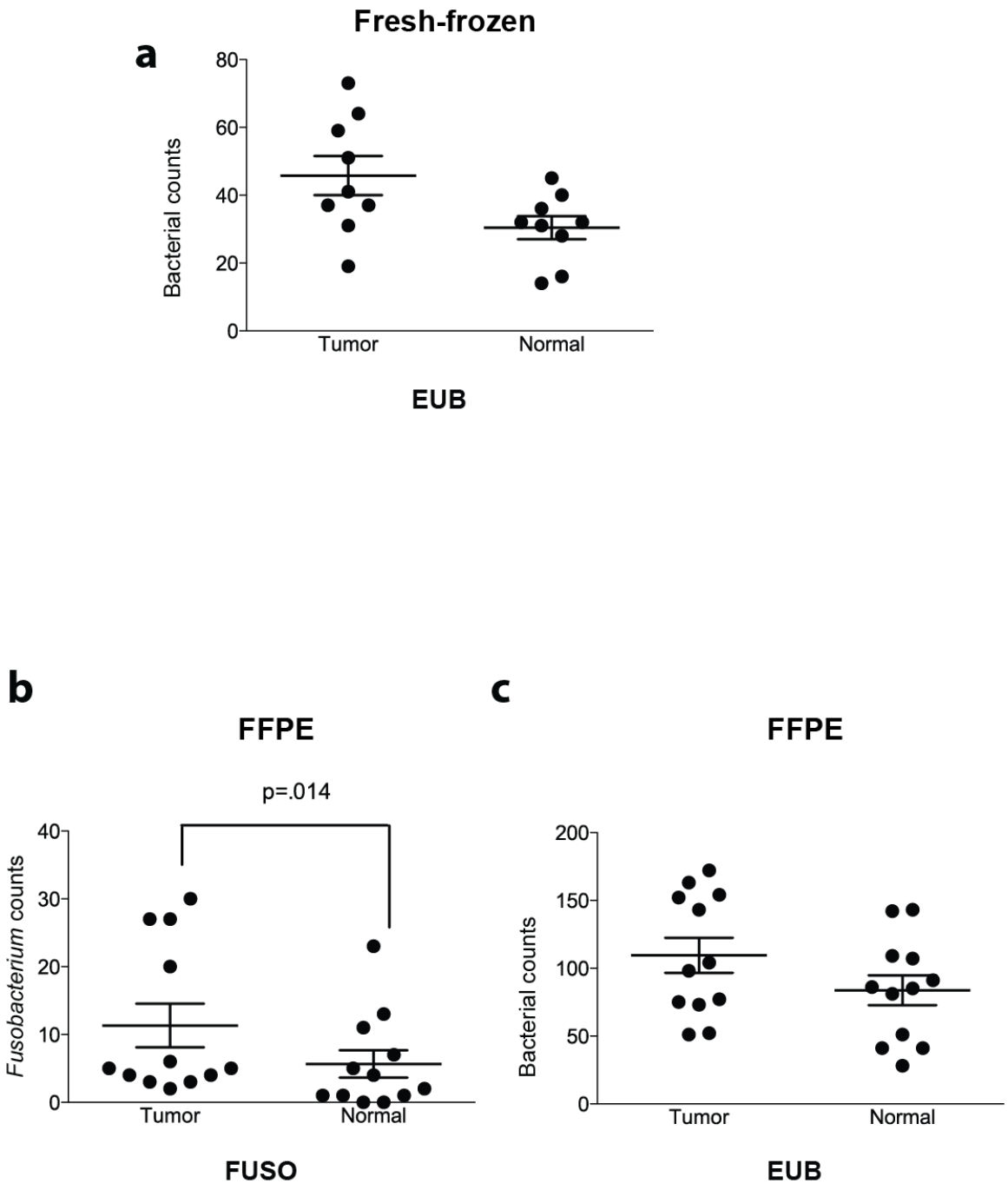
Supplementary Figure 2-4. Microbial diversity in the colon cancer microbiome. Comparison of the microbial diversity in tumor and normal tissues by rarefaction analysis on the OTUs detected by 16S rDNA sequencing **(a)** across all samples and **(b)** on a sample-by sample basis by using the QIIME package chao1 metric (see **Materials and Methods**), a species richness indicator. **(c)** Correlation of species richness with tumor stage by the chao1 metric (red: normal; blue: stage I; orange: stage II; green: stage III; purple: stage IV) demonstrates that higher stage tumors exhibit greater species richness.



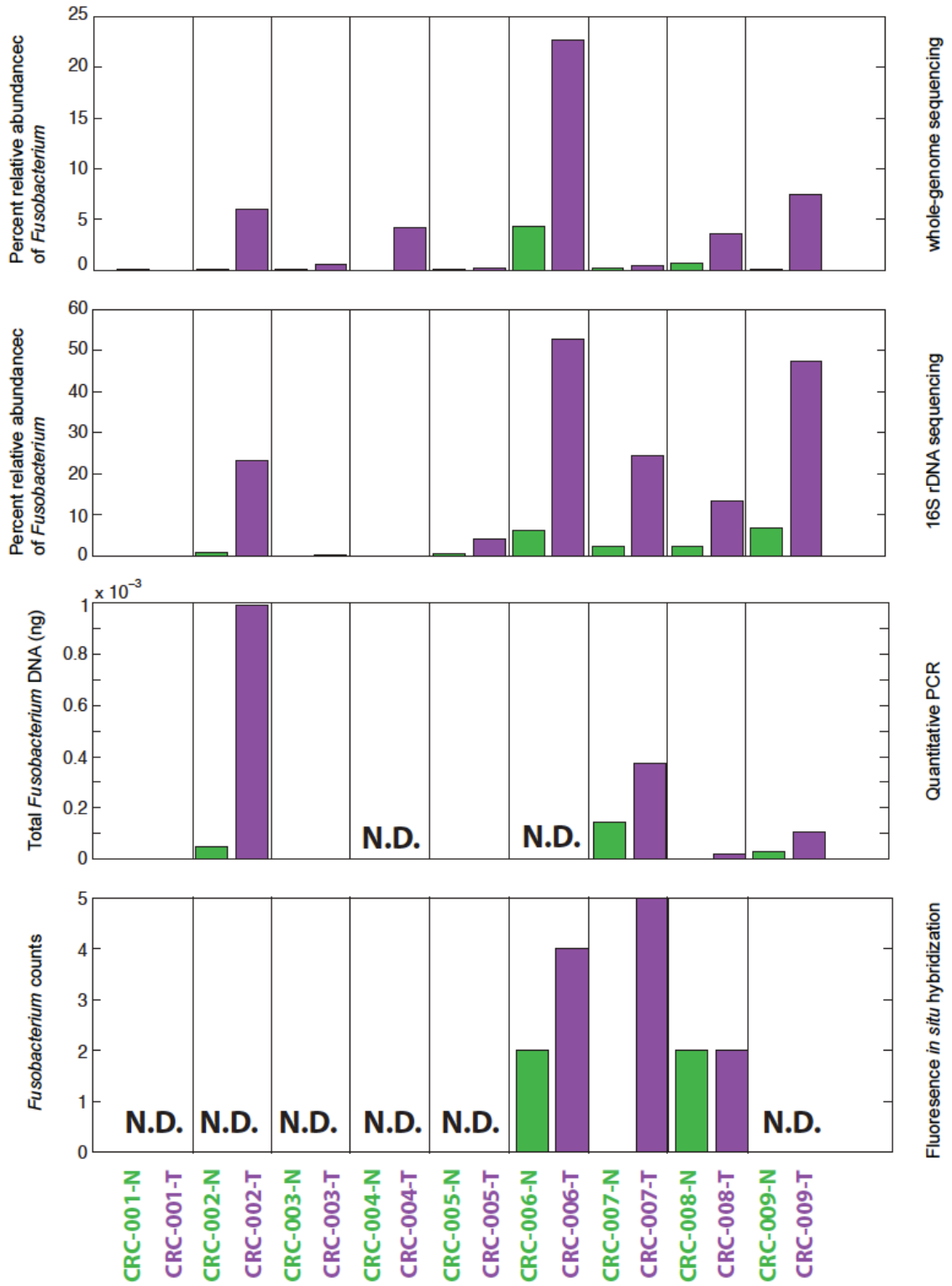
Supplementary Figure 2-5. *Fusobacterium* load correlates with geography. 16S rDNA sequencing shows an increased mean relative abundance of *Fusobacterium* OTUs in tumors from patients in Barcelona, Spain versus GCI samples, which are a composite collection consisting of samples from the United States and Vietnam. P-value calculated using a two-sample t-test.



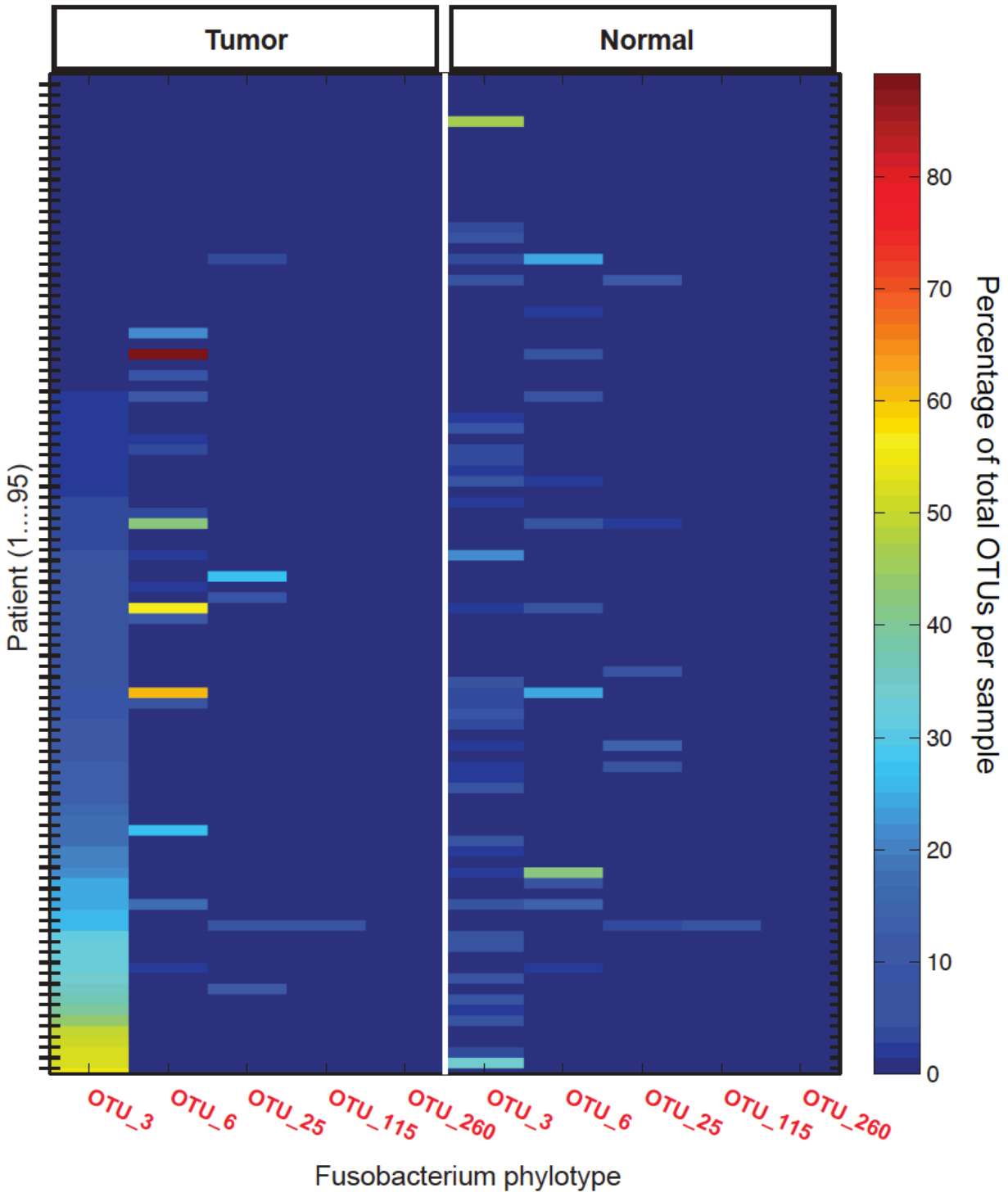
Supplementary Figure 2-6. Validation of the tumor-enrichment of *Fusobacterium* by quantitative PCR. **(a)** Absolute quantification of *Fusbacterium* DNA relative to a standard curve produced using *Fusobacterium nucleatum* genomic DNA and quantitative PCR probe specific to the *Fusobacterium* genus. 19 data points are beyond the y-axis limits but are reflected in the mean and standard error of the mean indicated on the plot. P-value calculated by a Wilcoxon matched-pairs signed rank test (non-parametric). **(b)** Quantification of Fusobacterium in tumor relative to normal by use of the ΔΔCt method quantitated against a probe targeting human endogenous 18S.



Supplementary Figure 2-7. Fluorescence *in situ* hybridization (FISH) detects enrichment of fusobacteria in colorectal tumors. (a) Total bacterial count data using an Oregon-Green 488-conjugated ''universal bacterial'' 16S rDNA-directed oligonucleotide probe (EUB338) is shown for 3 random 40X fields chosen for scoring by an observer blinded to tumor/normal status. Total bacterial counts were not enriched in either tumor or normals (P = 0.0517). This corresponds to Fig. 3B, which shows *Fusobacterium* counts. FISH was also performed on formalin-fixed, paraffin-embedded (FFPE) colorectal tumor and normal tissues using 5 random 40X fields with the Cy3- conjugated Fusobacterium-directed (FUSO) probe (enriched in tumor; p = 0.014). Epithelial cell nuclei were stained with DAPI **(b)** and the ''universal bacterial'' EUB probe (no enrichment; p = 0.1024) **(c)**. All p-values were calculated using a Wilcoxon matched-pairs signed rank test (non-parametric).

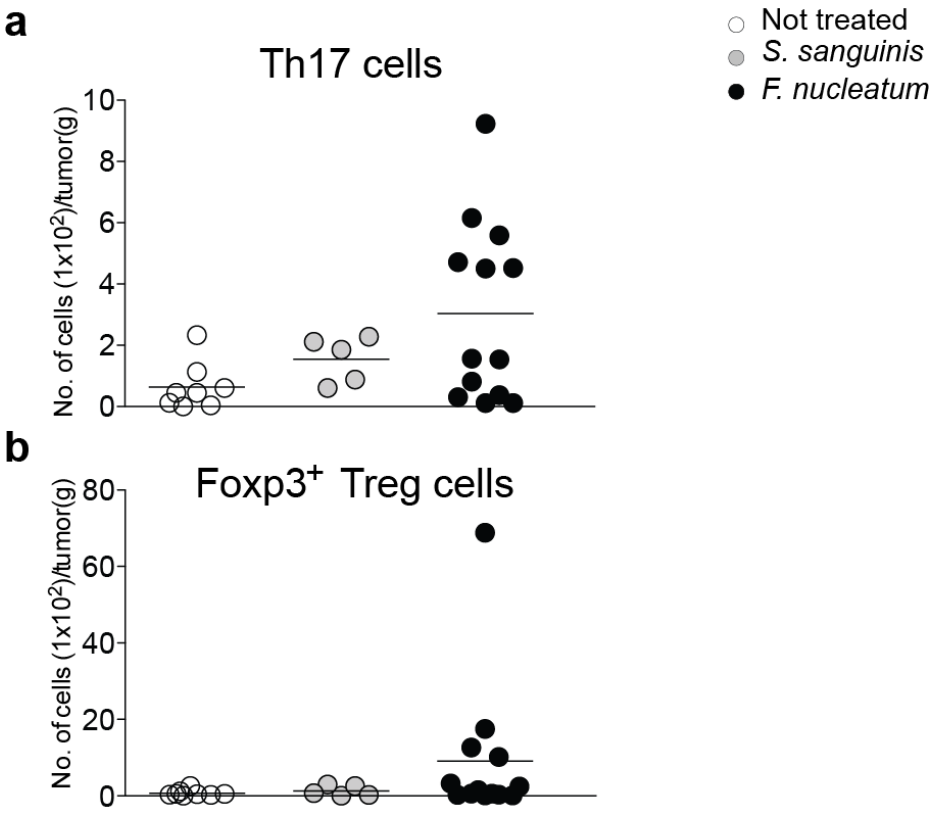


Supplementary Figure 2-8. Quantitation of *Fusobacterium* across analytic methods. A comparison of *Fusobacterium* quantitation is shown for the first 9 sample pairs by whole-genome sequencing relative abundance, 16S rDNA sequencing relative abundance, absolute *Fusobacterium* DNA levels by quantitative PCR, and Fusobacterium cell counts by fluorescence *in situ* hybridization. "N.D." indicates "not determined."

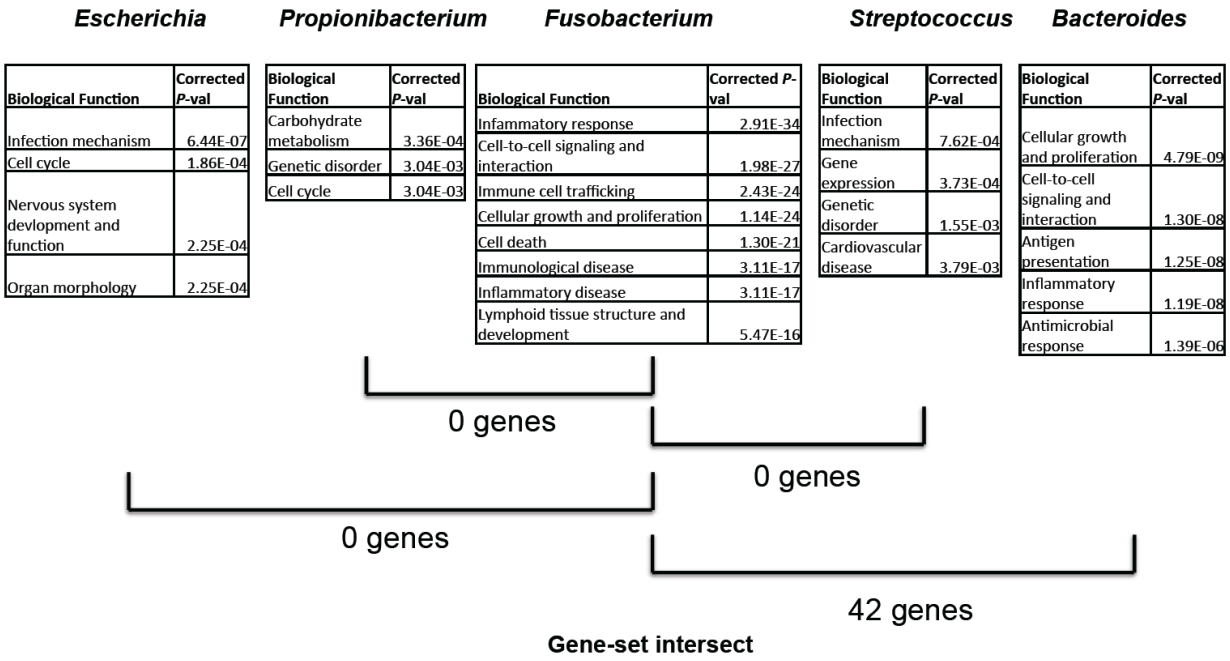


Supplementary Figure 2-9. Relative abundance measurements for *Fusobacterium* OTUs in human colon cancer tissues. The relative abundance of each of the five major OTUs in colon tumors (left panel) and normal tissue (right panel) is represented on a per-patient basis (each row represents a different individual). The percent relative abundance is denoted by a color gradient indicating relatively low abundance in blue (0-10%), very high abundance in dark red (>80%), and a range of other colors to specify intermediate abundance. The percentage signifies the abundance of the indicated OTU relative to all other phylotypes in

that specimen.



Supplementary Figure 3-1. *F. nucleatum* does not affect intratumoral Th17 nor Foxp3$^+$ Treg cell numbers. Flow cytometric analyses: Cell number/gram tumor of (**a**) CD3$^+$CD4$^+$ IL-17$^+$ T cells and (**b**) CD3$^+$CD4$^+$ Foxp3$^+$ T cells. Each symbol represents data from an individual mouse.



Supplementary Figure 3-2. Human colon tumor RNA-Seq analysis shows enrichment for inflammatory functions in genes correlated with *Fusobacterium* abundance but not the abundance of other genera. Spearman's rank correlations of genus-level microbial relative abundance with host gene expression was used to weight Gene Ontology enrichment using Ingenuity Pathway Analysis software for the top 5 most abundant genera. Gene-set intersect indicates the number of genes that are shared in the enriched gene sets between *Fusobacterium* and each of the other genera.

Supplementary Table 1-1. Bacterial genomes used to construct an artificial sequence dataset to test the metagenomics module of PathSeq. Shown is the set of bacterial genomes that were each used to create a set of 10,000 random 100-mers. The species were chosen to represent both evolutionary relatedness and divergence to test the ability of the metagenomics module to properly assess the microbial representation of a mixed sample.

Accession	Definition
NC_003228.3	Bacteroides fragilis NCTC 9343, complete genome.
NC_009614.1	Bacteroides vulgatus ATCC 8482, complete genome.
NC_013316.1	Clostridium difficile R20291, complete genome.
NC_009615.1	Parabacteroides distasonis ATCC 8503, complete genome.
NC_004663.1	Bacteroides thetaiotaomicron VPI-5482, complete genome.
NC_010816.1	Bifidobacterium longum DJO10A, complete genome.
NC_008618.1	Bifidobacterium adolescentis ATCC 15703, complete genome.
NC_012781.1	Eubacterium rectale ATCC 33656, complete genome.
NC_012778.1	Eubacterium eligens ATCC 27750, complete genome.
NC_010655.1	Akkermansia muciniphila ATCC BAA-835, complete genome.
NC_011353.1	Escherichia coli O157:H7 str. EC4115, complete genome.
NC_004668.1	Enterococcus faecalis V583, complete genome.

Supplementary Table 1-2. Metagenomic analysis on a sequence dataset constructed from a set of twelve bacterial genomes. Shown is the number of reads that were identified as matching to the indicated bacterial genome (actual number of reads is 10,000 for each species).

Genus	Species	Number of Reads	Fraction Genome Coverage
Bifidobacterium	adolescentis	9877	0.376091154
Eubacterium	eligens	9903	0.368491132
Bifidobacterium	longum	9933	0.333497209
Akkermansia	muciniphila	9954	0.310035802
Enterococcus	faecalis	9900	0.264391797
Eubacterium	rectale	9953	0.249987753
Clostridium	difficile	9829	0.208571056
Parabacteroides	distasonis	9929	0.186090932
Bacteroides	vulgatus	9943	0.173829585
Bacteroides	fragilis	9929	0.169107843
Bacteroides	thetaiotaomicron	9933	0.146473822
Escherichia	coli	5133	0.089648584
Staphylococcus	aureus	3	1.04E-04
Bifidobacterium	dentium	1	3.79E-05
Shigella	dysenteriae	1	2.29E-05
Shigella	sonnei	1	2.07E-05

Supplementary Table 1-3. PathSeq performance on artificially generated sequence data. Reads were generated by sampling random 100-mer sequences from a human transcriptome database, generating 20 million reads, or from a set of twelve virus genomes each substitutionally mutated at twelve distinct rates, generating 144,000 reads. The rows represent the number of reads remaining at each step in the PathSeq workflow.

Stage in workflow	Human reads remaining	Virus reads remaining	Virus reads subtracted at each step
START	20000000	144000	
Duplicate Remover	19633552	144000	0
Maq database1	5441980	144000	0
Maq database2	696074	144000	0
Maq database3	1218	144000	0
Maq database4	1213	144000	0
RepeatMasker Remover	853	142880	1120
MegaBlast database1	0	142878	2
MegaBlast database2	0	142878	0
Blast database1	0	142878	0
Blast database2	0	142878	0

Supplementary Table 1-4. PathSeq performance on artificially generated sequence data with introduced sequencing errors. Reads are the same as in **Supplementary Table 1-3** except that "sequence errors" were introduced into the reads according to a sequencing error distribution model.

Stage in workflow	Human reads remaining	Virus reads remaining	Virus reads subtracted at each step
START	20000000	144000	
Duplicate Remover	19999188	144000	0
Maq database1	14240412	144000	0
Maq database2	12205718	144000	0
Maq database3	11745641	144000	0
Maq database4	11744303	144000	0
RepeatMasker Remover	11149237	143071	929
MegaBlast database1	54343	143070	1
MegaBlast database2	54329	143070	0
Blast database1	0	143069	1
Blast database2	0	143069	0

Supplementary Table 2-1. PathSeq run-statistics and microbial classifications using colon cancer human whole-genome sequencing data. Shown is the total number of reads, total number of quality-filtered (QF) reads, and reads identified as microbial sequences following PathSeq analysis (see **Materials and Methods**) for each of the 9 colorectal tumor/normal pairs.

Sample	Total Reads	Total QF Reads	Microbial reads
CRC-0001-N	1502889378	53338560	1079732
CRC-0001-T	1498449436	48610083	127532
CRC-0002-N	1526909584	302588252	111186
CRC-0002-T	1472368112	253126272	237632
CRC-0003-N	2292261110	74512887	31346
CRC-0003-T	1986594024	72036170	4153
CRC-0004-N	1314312300	212000000	701
CRC-0004-T	1384860904	65391426	2242
CRC-0005-N	1509814586	8000000	26020
CRC-0005-T	2008090000	20000000	3806
CRC-0006-N	1634490198	12000000	146061
CRC-0006-T	1586878006	34716131	179982
CRC-0007-N	1538750552	72374541	138079
CRC-0007-T	1389724114	58661960	90406
CRC-0008-N	1557712266	65288952	7916
CRC-0008-T	1449651044	257878732	16429
CRC-0009-N	1224553004	175370564	344055
CRC-0009-T	1252693302	167829028	269547
MEAN	1562833440	108540197.7	156490.2778
MEDIAN	1506351982	68713798	100796

Supplementary Table 2-2. Identification of viruses in colon cancer human whole-genome sequencing data by PathSeq analysis. Virus sequences were identified among the remaining non-human reads following analysis with PathSeq.

Sample	Virus name (Number of reads matching virus sequence)
CRC-0001-N	Norwalk virus(7), parvovirus B19(2)
CRC-0001-T	Human herpesvirus 4 (2)
CRC-0002-N	Human herpesvirus 7 (8), Human herpesvirus 4(1), parvovirus B19(5)
CRC-0002-T	Human herpesvirus 5 (16), parvovirus B19 (3), Torque teno virus (3)
CRC-0003-N	Human herpesvirus 6 serotype B (333), Human herpesvirus 7 (14)
CRC-0003-T	Human herpesvirus 6 serotype B (1)
CRC-0004-N	Human herpesvirus 7(2)
CRC-0004-T	None detected
CRC-0005-N	Human herpesvirus 7 (80), Human herpesvirus 5 (5), Human herpesvirus 6 serotype B (1)
CRC-0005-T	Human herpesvirus 7 (1), Human herpesvirus 5 (3), Human herpesvirus 4 (1)
CRC-0006-N	Human herpesvirus 7 (3)
CRC-0006-T	None detected
CRC-0007-N	Human herpesvirus 7 (6), Human herpesvirus 6 serotype B (6), Human herpesvirus 4 (3)
CRC-0007-T	Polyomavirus HPyV6 (1)
CRC-0008-N	Human herpesvirus 7 (2)
CRC-0008-T	Human herpesvirus 4 (10), Xenotropic MuLV-related virus (3)
CRC-0009-N	Human herpesvirus 7 (52), Human herpesvirus 6 serotype B (94)
CRC-0009-T	Human herpesvirus 7 (105), Human herpesvirus 6 serotype B (216)

Supplementary Table 2-3. Quantitative PCR for Fusobacterium in colon cancer cell lines. Quantitative PCR was performed using a *Fusobacterium*-directed probe compared against a human endogenous 18S-directed probe on both DNA and RNA extracted from each of the indicated colon cancer cell lines. The qPCR signal was strong for the endogenous control target but completely absent for the fusobacterial target. All experiments were performed in triplicate.

Cell Line ID	
CL-40	CL-11
MDST8	HT115
Hs 698.T	T84
LS123	SNU-C5
C2BBe1	SW1463
SNU-283	HCT-15_DLD-1
Hs 675.T	NCI-H747
COLO-320	LS1034
CL-14	SNU-81
HCC-56	HCT-8
COLO-678	SW1417
CCK-81	SNU-C4
SNU-503	LS513
HCT 116	SK-CO-1
SNU-61	LS 180
SNU-C2A	HT-29
SW948	HT115
SW837	SNU-1040
LoVo	RKO
SNU-C1	HT55
GP2d	NCI-H508
CL-40	COLO 741
NCI-H716	LS411N
CL-34	HuTu 80
HCT-15_DLD-1	KM12
CL-14	SNU-407
MDST8	SW48
OUMS-23	SNU-175
SNU-1033	HT55
SW403	

Supplementary Table 2-4. Quantitative PCR for *Fusobacterium* in metastases from colorectal carcinoma. Quantitative PCR using a probe targeting a region of 16S rDNA conserved among fusobacteria (see Methods) shows that 2 of 11 colon metastases to the liver or lymph nodes have a positive signal. All positive results were positive in three replicates.

Sample	Sample Type	Task	Detector	Avg Ct
0021M	metastasis	Endogenous Control	Endogenous 18S	20.99
0021M	metastasis	Target	Fuso probe	40
0048M	metastasis	Endogenous Control	Endogenous 18S	22.719
0048M	metastasis	Target	Fuso probe	40
0070M	metastasis	Endogenous Control	Endogenous 18S	21.478
0070M	metastasis	Target	Fuso probe	40
0079M	metastasis	Endogenous Control	Endogenous 18S	21.85
0079M	metastasis	Target	Fuso probe	**35.833**
0102M	metastasis	Endogenous Control	Endogenous 18S	19.872
0102M	metastasis	Target	Fuso probe	40
0111M	metastasis	Endogenous Control	Endogenous 18S	16.769
0111M	metastasis	Target	Fuso probe	40
0112M	metastasis	Endogenous Control	Endogenous 18S	18.909
0112M	metastasis	Target	Fuso probe	40
0114M	metastasis	Endogenous Control	Endogenous 18S	22.461
0114M	metastasis	Target	Fuso probe	**36.629**
0120M	metastasis	Endogenous Control	Endogenous 18S	20.834
0120M	metastasis	Target	Fuso probe	40
0121M	metastasis	Endogenous Control	Endogenous 18S	23.248
0121M	metastasis	Target	Fuso probe	40
0123M	metastasis	Endogenous Control	Endogenous 18S	18.225
0123M	metastasis	Target	Fuso probe	40
0131M	metastasis	Endogenous Control	Endogenous 18S	22.002
0131M	metastasis	Target	Fuso probe	40
090P	primary	Endogenous Control	Endogenous 18S	18.015
090P	primary	Target	Fuso probe	40

Supplementary Table 2-5. Reference strain identification for Fusobacterium phylogenetic analysis. Abbreviated names used in the phylogenetic tree in Fig. 2-4a are specified using complete Greengenes strain names.

Designation in Fig. 4A	Greengenes strain name
F. naviforme	F. naviforme
F. russii str. 1	F. russii str. ATCC 25533T
F. russii	F. russii
F. simiae	F. simiae
F. nucleatum str. 1	F. nucleatum fusiforme subsp. fusforme str. NCTC F. russii T
F. simiae str. 1	F. simiae str. ATCC 33568T
F. genomosp. C2	F. genomosp. C2
F. genomosp. C1	F. genomosp. C1
F. nucleatum str. 2	F. ATCC 25586 subsp. nucleatum str. JCM 8532
F. nucleatum	F. subsp. nucleatum str. ATCC25586
F. naviforme str. 1	F. naviforme str. DMS 20699 NCTC 11464
F. nucleatum str. 3	F. nucleatum subsp. vincentii str. ATCC 49256
F. sp. str. FL002	F. sp. oral taxon 205 str. FL002
F. nucleatum str. 4	F. nucleatum str. JC-208
F. nucleatum str. 5	F. nucleatum ATCC 10953 subsp. polymorphum str. JCM 12990
F. canifelinum	F. canifelinum str. RMA 12708
F. periodonticum	F. periodonticum str. KP-F10
F. perfoetens	F. perfoetens
F. necrogenes	F. necrogenes str. ATCC 25556T
F. mortiferum	F. mortiferum
F. varium str. 1	F. varium str. NCTC 10560T
F. mortiferum str. 1	F. mortiferum str. ATCC 25557T
F. varium	F. varium
F. ulcerans str. 1	F. ulcerans str. NCTC 12111T
F. equinum str. 1	F. equinum str. horse VPB 4027
F. gonidoformans	F. gonidoformans
F. necrophorum str. 1	F. necrophorum str. AB FnS-1
F. sp. str. D12	F. sp. str. D12
F. necrophorum	F. necrophorum
F. necrophorum str. 2	F. necrophorum subsp. funduliforme str. DSM 19678
F. gonidiaformans	F. gonidiaformans str. ATCC 25563T

Supplementary Table 3-1. Sample list of all colonic adenomas assessed for *Fusobacterium* abundance.

Sample	Origin	ΔCt (Fuso - Eubact)
n01	U. of Pennsylvania	Fuso (-)
a01	U. of Pennsylvania	Fuso (-)
n02	U. of Virginia Health System	Fuso (-)
a02	U. of Virginia Health System	Fuso (-)
n03	U. of Virginia Health System	Fuso (-)
a03	U. of Virginia Health System	17.36
n04	U. of Virginia Health System	Fuso (-)
a04	U. of Virginia Health System	Fuso (-)
n05	U. of Virginia Health System	Fuso (-)
a05	U. of Virginia Health System	Fuso (-)
n06	Mass. Gen. Hosp.	18.29
a06	Mass. Gen. Hosp.	17.65
n07	Mass. Gen. Hosp.	Fuso (-)
a07	Mass. Gen. Hosp.	Fuso (-)
n08	Mass. Gen. Hosp.	Fuso (-)
a08	Mass. Gen. Hosp.	Fuso (-)
n09	Mass. Gen. Hosp.	Fuso (-)
a09	Mass. Gen. Hosp.	20.35
n10	Vanderbilt U. Med. Center	26.62
a10	Vanderbilt U. Med. Center	15.35
n11	Vanderbilt U. Med. Center	Fuso (-)
a11	Vanderbilt U. Med. Center	12.29
n12	Vanderbilt U. Med. Center	14.41
a12	Vanderbilt U. Med. Center	14.08
n13	Vanderbilt U. Med. Center	Fuso (-)
a13	Vanderbilt U. Med. Center	Fuso (-)
n14	Vanderbilt U. Med. Center	19.42
a14	Vanderbilt U. Med. Center	10.75
n15	Vanderbilt U. Med. Center	Fuso (-)
a15	Vanderbilt U. Med. Center	Fuso (-)
n16	U. of Aberdeen	Fuso (-)
a16	U. of Aberdeen	Fuso (-)
n17	U. of Aberdeen	17.96
a17	U. of Aberdeen	14.43
n18	U. of Aberdeen	17.97
a18	U. of Aberdeen	17.04
n19	U. of Aberdeen	Fuso (-)
a19	U. of Aberdeen	21.30
n20	U. of Aberdeen	11.44
a20	U. of Aberdeen	11.59
n21	U. of Aberdeen	Fuso (-)
a21	U. of Aberdeen	Fuso (-)
n22	U. of Aberdeen	25.53
a22	U. of Aberdeen	23.98
n23	U. of Aberdeen	Fuso (-)
a23	U. of Aberdeen	Fuso (-)
n24	U. of Aberdeen	21.34
a24	U. of Aberdeen	Fuso (-)
n25	U. of Aberdeen	24.19
a25	U. of Aberdeen	21.91
n26	U. of Aberdeen	21.36
a26	U. of Aberdeen	20.86
n27	U. of Aberdeen	22.72
a27	U. of Aberdeen	18.91
n28	U. of Aberdeen	Fuso (-)
a28	U. of Aberdeen	Fuso (-)
n29	U. of Aberdeen	Fuso (-)
a29	U. of Aberdeen	Fuso (-)
n30	U. of Aberdeen	Fuso (-)
a30	U. of Aberdeen	Fuso (-)
n31	U. of Aberdeen	Fuso (-)
a31	U. of Aberdeen	Fuso (-)

Supplementary Table 3-2. *Fusobacterium* is detected at a higher abundance in stool from CRC and adenoma cases than from healthy controls.

	# indiv.	Fuso (-)	Fuso (+)	*P* value
Normal	31	16	15	
Adenoma	28	4	24	$P < 0.005$
Carcinoma	27	0	27	$P < 0.00005$